Representativeness of stations and reliability of data in the Swiss Phenology Network : Technical Report MeteoSwiss No. 267

Güsewell, Sabine ; Pietragalla, Barbara ; Gehrig, Regula ; Furrer, Reinhard

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Home Affairs FDHA
Federal Office of Meteorology and Climatology MeteoSwiss

MeteoSwiss

Technical Report MeteoSwiss No. 267

Representativeness of stations and reliability of data in the Swiss Phenology Network

Sabine Güsewell, Barbara Pietragalla, Regula Gehrig and Reinhard Furrer

ISSN: 2296-0058

Technical Report MeteoSwiss No. 267

Representativeness of stations and reliability of data in the Swiss Phenology Network

Sabine Güsewell[1] , Barbara Pietragalla, Regula Gehrig and Reinhard Furrer[2]

[1] Plant Ecology, Institute of Integrative Biology, ETH Zurich, 8092 Zurich

[2] Department of Mathematics and Department of Computational Science, University of Zurich, 8057 Zurich

Recommended citation:
Güsewell S, Pietragalla B, Gehrig R and Furrer R: 2018, Representativeness of stations and reliability of data in the Swiss Phenology Network, *Technical Report MeteoSwiss*, 267, 100 pp.

Editor:
Federal Office of Meteorology and Climatology, MeteoSwiss, © 2018

MeteoSwiss
Operation Center 1
CH-8058 Zürich-Flughafen
T +41 58 460 91 11
www.meteoswiss.ch

Abstract

Phenological observation networks have been implemented in many countries to monitor how the timing of plant seasonal life cycles varies in space and time. Data are used to model the responses of plant phenology to climatic factors and to predict changes associated with future climate warming. The quality of these predictions depends critically on the quality of the underlying data.

This study evaluates the representativeness and precision of data from the Swiss Phenology Network, which was implemented in 1951 by MeteoSwiss, and which currently includes 167 stations and 69 phenophases. The onset dates of these phenophases are recorded annually at each station by volunteering observers, leading to a data set with 186171 observations between 1951 and 2012. We analyse the spatial structure of phenological variation (considering mean onset dates through time, between-year variation and long-term trends), phenological responses to temperature through time and space, similarities in phenological time series between stations, and their use in predictive models for error detection.

Results show that phenological variation across Switzerland is determined by altitude, large-scale spatial trends and local deviations (e.g. due to variation among individual plants and observation error), whereas small-scale spatial dependence (correlation of neighbouring stations) is weak. The number of stations currently included in the Swiss Phenology Network is sufficient for precise estimates of mean onset dates of each phenophase, of long-term trends and of responses to temperature for the entire country and for three altitudinal layers. More stations would be needed in some regions for a precise analysis of regional differences. The network does currently not include groups of stations with similar patterns of between-year variation for all phenological stages, i.e. no redundancy. A comparison of predictive models suggests that models with additive random effects of station and year (or station and year-specific temperature) are most suitable for data quality checking in practice. The inclusion of data from neighbouring stations, for other phenophases, or from the previous year hardly improves the detection of erroneous data entries.

We conclude that the precision of results obtained from the Swiss Phenology Network depends more on the number of stations included in the network than on their exact geographic distribution as long as all regions are sufficiently represented. Given the important effect of altitude on phenological variation, the availability of phenological stations over a broad altitudinal range is a particular asset of the Swiss Phenology Network, which should be maintained.

Zusammenfassung

Phänologische Beobachtungsnetze wurden in vielen Ländern eingerichtet, um Veränderungen der jährlichen Lebenszyklen der Pflanzen in Raum und Zeit aufzuzeichnen. Die Daten werden dazu verwendet, die pflanzenphänologischen Reaktionen auf klimatische Faktoren zu modellieren und phänologische Veränderungen infolge der zukünftigen Klimaerwärmung vorherzusagen. Die Qualität dieser Vorhersagen hängt entscheidend von der Qualität der erfassten Daten ab.

Diese Arbeit untersucht die Repräsentativität und Genauigkeit der Daten des Schweizer Phänologie-Beobachtungsnetzes, welches 1951 von MeteoSchweiz gegründet wurde und derzeit 167 Stationen und 69 Phänophasen umfasst. Die Eintrittstermine dieser Phänophasen werden jährlich von freiwilligen Beobachtern notiert, was von 1951 bis 2012 einen Datensatz mit 186171 Beobachtungen ergab. Wir analysieren die räumliche Struktur der phänologischen Unterschiede (unter Berücksichtigung der zeitlichen Variation der Eintrittstermine, der Variation von Jahr zu Jahr und der Langzeittrends), die phänologischen Reaktionen auf Temperaturunterschiede sowie Ähnlichkeiten phänologischer Datenreihen von verschiedenen Stationen und ihre Verwendung in Prognosemodellen zur Fehlererkennung.

Die Ergebnisse zeigen, dass die phänologischen Unterschiede in der Schweiz durch die Höhenlage, durch grossräumige Trends und lokale Abweichungen (Unterschiede zwischen einzelnen Pflanzen sowie Beobachtungsfehler) bestimmt werden, während räumliche Abhängigkeiten (Korrelation benachbarter Stationen) nur schwach sind. Die Anzahl der aktuell zum Schweizer Phänologie-Beobachtungsnetz gehörenden Stationen genügt für die genaue Bestimmung der mittleren Eintrittstermine aller Phänophasen, von Langzeittrends und von Reaktionen auf die Temperatur mit Bezug auf das ganze Land oder drei Höhenstufen. Mehr Stationen wären in manchen Regionen für eine genaue Analyse regionaler Unterschiede nötig. Das Netz enthält derzeit keine Stationen, die für alle Phänophasen das gleiche Muster der Jahr-zu-Jahr-Variation aufweisen, es gibt also keine Redundanz. Der Vergleich von Prognosemodellen für Eintrittstermine einzelner Phänophasen deutet an, dass Modelle mit unabhängigen Zufallseffekten der Station und des Jahres (oder der Station und Temperaturabweichung vom Langzeitmittel) für eine Datenqualitätskontrolle am besten geeignet sind. Das Einbeziehen der Daten von Nachbarstationen, anderer Phänophasen oder aus dem Vorjahr verbessert nicht das Erkennen fehlerhafter Daten.

Die Genauigkeit der Ergebnisse, die mit Daten des Schweizer Phänologie-Beobachtungsnetzes erzielt werden, hängt mehr von der Anzahl der vorhandenen Stationen ab als von deren genauen geographischen Verteilung, vorausgesetzt alle Regionen sind genügend berücksichtigt. Wegen dem wichtigen Einfluss der Höhenlage auf die phänologischen Unterschiede, ist das Vorhandensein von phänologischen Stationen mit einer weiten Spanne unterschiedlicher Höhenlagen ein besonderer Wert des Schweizer Phänologie-Beobachtungsnetzes, den es zu erhalten gilt.

Contents

1 Introduction

1.1 Background

Phenological shifts - changes in the timing of seasonal life-cycle events - are currently the most conspicuous response of living organisms to climate warming (*Parmesan*, 2007; *Cleland et al.*, 2007). They document unequivocally the effects of climate change on individuals and communities, and they are expected to be critical for the ability of plant and animal species to adjust to new conditions (*Both et al.*, 2009; *Cleland et al.*, 2012; *Miller-Rushing et al.*, 2010). Therefore, phenological studies are central for our understanding of the ecological implications of climate change (*Morisette et al.*, 2009).

While early research focused on demonstrating the generality of phenological shifts due to recent climate change, there has rapidly been increasing interest in the variability of observed shifts (*Defila and Clot*, 2001; *Menzel et al.*, 2006; *Parmesan*, 2007; *Primack et al.*, 2009; *Schleip et al.*, 2009). Shifts were found to differ among phenological phases (*Menzel et al.*, 2001), among plant species and functional types (*Fitter and Fitter*, 2002; *Ziello et al.*, 2012), among regions (*Menzel et al.*, 2006; *Studer et al.*, 2005), between urban and rural areas (*Roetzer et al.*, 2000), and among individual stations, even at a small scale (*Menzel et al.*, 2001). These multiple sources of variation limit our ability to predict whether and how phenological changes will continue in the future (*Dose and Menzel*, 2004; *Cleland et al.*, 2007).

Phenological observation networks are an essential source of information about recent changes in plant phenology (*Koch*, 2010; *Schwartz*, 2013). Phenological networks consist of multiple stations within a region, where observers (usually volunteers) record the timing of particular phenophases every year according to standardized protocols. Data are collected and managed by a central organisation, usually associated with national meteorological institutes (*Menzel*, 2013). Some of these networks have been established many decades ago to characterize regional and local bioclimates and to inform agriculture, horticulture and forestry (*Chmielewski*, 2013). The spatial and temporal coverage of the resulting datasets makes them extremely valuable for climate change research (*Dierenbach et al.*, 2013). These assets outweigh some inevitable drawbacks, such as unequal distribution of sampling stations, limited precision of observations, and discontinuity or inconsistencies in data collection (*Beaubien and Hamann*, 2011). However, the analysis of phenological network data typically requires an intensive preliminary data cleaning process with partly subjective decisions on the validity of certain observations (e.g. *Menzel et al.*, 2001).

The continued importance of phenological networks for climate change research calls for increased efforts to improve the accuracy of data and the representativeness of stations included in these networks (*Schaber and Badeck*, 2002). Data accuracy depends on precise observation protocols (*Brügger and*

Vassella, 2003), diligent observers and a data validation procedure to eliminate deviant, probably erroneous records (*Schaber and Badeck*, 2002; *Schaber et al.*, 2010; *Beaubien and Hamann*, 2011). However, data management and quality tests are costly (*Schwartz et al.*, 2013), so that there is a need to balance the gain in precision achieved by more detailed data checking against the associated additional costs. As data are increasingly supplied by observers in electronic form, automatic procedures for error detection are becoming the main tool for data quality checks.

Representativeness is a key requirement for climatological monitoring systems in general. This concept includes several aspects. Observations should adequately represent the full range of spatial and temporal variation in the studied phenomena. This requires a sufficient number and an appropriate spatial distribution of stations (*Gehrig*, 2012). The number of stations should also allow a sufficiently precise estimation of relevant parameters. If a pre-existing classification of stations is expected to be relevant for the studied phenomena (e.g. topographic regions), each class should be represented with a sufficient number of stations for reliable group-level estimates. 'Representativeness' can further refer to the fact that some stations are sufficiently similar to each other that groups of similar stations could be represented adequately by one of them (*Begert*, 2008; *DeGaetano*, 2001). The identification of these representative stations may help in setting priorities among stations if a choice has to be made (*DeGaetano*, 2001). Furthermore, similarities between stations and between variables can help in assessing the reliability of data and the consistency of results derived from them (*Linkosalo et al.*, 1996; *Linkosalo*, 2000; *Gehrig*, 2012). Similarities between neighbouring stations (spatial correlation) can be exploited for spatial smoothing, e.g. to construct phenological maps (*Schleip et al.*, 2009). However, strong similarities between stations (especially between neighbouring ones) may imply that some stations provide redundant information and do not fully contribute to the representativeness of the network.

Statistical approaches to assess representativeness include correlation and cluster analyses as well as regression models in a broad sense. Correlations are a simple way to describe similarities between stations in a multidimensional dataset. Cluster analysis identifies groups of stations with similar values or with similar temporal patterns for the variables of interest. The resulting groups can be the basis for summarizing the data in a meaningful way (*Begert*, 2008). If groups are contiguous in space, they delimit homogeneous regions with respect to the climatic parameters considered (*Begert*, 2008; *DeGaetano*, 2001), so that predictions can also be made for other locations within the same region (*Gehrig*, 2012). Finally, regression methods and extensions such as mixed models and spatial models relate variation in the observed data to pre-existing classifications, relevant drivers (e.g. temperature), time and space. A comprehensive assessment of representativeness therefore requires a combination of different statistical methods.

1.2 Aims and structure of the report

The present report evaluates the representativeness of the Swiss Phenology Network, which was implemented in 1951 by MeteoSwiss. So far, representativeness has been evaluated for other climatological monitoring systems of MeteoSwiss, including the Swiss National Basic Climatological Network (*Begert*, 2008), measurements of snow cover (*Wüthrich et al.*, 2010) and the pollen monitoring network (*Gehrig*, 2012). Data from the Swiss Phenology Network have been used for comprehensive analyses of spatial and temporal trends (*Defila and Clot*, 2001; *Studer et al.*, 2005, 2007; *Menzel et al.*, 2006),

but its representativeness has not yet been studied systematically. Here we analyse the spatial and temporal variation of phenological onset dates as well as similarities between phenological stations in order to evaluate their representativeness. The stations of the Swiss Phenology Network and the analysed data are presented in Chapter 2.

Spatial variation occurs in three dimensions and at different scales, raising the question which aspect of spatial variation should be represented by the Network. Previous analyses have highlighted the effect of altitude on means and trends of phenological onset dates (*Defila and Clot*, 2001; *Ziello et al.*, 2009). Furthermore, *Studer et al.* (2005) found regional differences in phenological trends related to precipitation and continentality. These results indicate that both altitude and the established subdivision of Switzerland into climatic regions could define relevant groups of stations for phenological monitoring. In addition, because the main driver of plant phenology, temperature, is spatially correlated (*Furrer and Sain*, 2009; *DeGaetano*, 2001), means and trends of phenological data may be spatially correlated as well. Based on these expectations, Chapter 3 investigates how phenological variation in Switzerland depends on altitude, climatic regions and spatial distance, and how the precision of parameter estimates depends on the number and distribution of stations with respect to these factors.

Besides spatial factors, ecological factors that vary at a local scale, such as genetic differences between and within plant populations, soil types, sunlight exposure, herbivores, diseases and human activities, may also influence plant phenology and its response to climate. As a result, the most similar phenological stations are not necessarily next to each other. Chapter 4 analyses similarities between stations with correlation and clustering methods to see if a simple and consistent classification of stations emerges, to identify the most representative stations, and to judge whether similarities between stations lead to redundancy in the Network.

The effect of temperature on plant phenology and its spatial variation across Switzerland has been evaluated in a multivariate approach by *Studer et al.* (2005) and as part of large-scale analyses by *Menzel et al.* (2006) and *Ziello et al.* (2009), while *Rutishauser et al.* (2008) analysed reconstructed time series back to 1702. Previous studies were generally based on time series until 2000 (except for *Rutishauser et al.*, 2008). The decade 1990–2000 was characterized by consistently high temperatures following a sudden warming in the late 1980ies. In contrast, the years after 2000 are characterized by strong fluctuations between extremely warm and 'normal' years. Hot spells are predicted to become more frequent in the future (*IPCC*, 2012), raising the question whether such events influence plant phenology differently from long-term trends or spatial gradients in temperature. With respect to the representativeness of the Swiss Phenology Network, the main question is how precisely and reliably phenological responses to temperature can be estimated with the available stations. These questions are addressed in Chapter 5.

To ensure data quality, new observations entered online in the Swiss Phenology Network are checked through an automatic plausibility test during data entry by the observer. This test is based on 95% prediction intervals (mean $\pm$ 2 sd) for each phenophase in each of five altitudinal layers; data entries that fall outside the prediction intervals generate a warning message inviting the observer to check his record. However, the prediction intervals are wide (on average 60 days), so that many errors cannot be detected without additional plausibility checks. More precise models would define smaller prediction intervals, so that errors could be detected more reliably. In a phenology network with multiple stations and variables, predictions could possibly be improved by including information from other phenophases or stations in the models (*Linkosalo*, 2000). This possibility is explored in Chapter 6 by comparing the

fit and predictive power of models with different predictors. The main conclusions are summarized in Chapter 7.

This study was initially part of a master thesis in Biostatistics at the University of Zurich (*Güsewell*, 2014). Because the results proved to be important for the design, further development and management of the Swiss Phenology Network, as well as for the climatological analysis of phenological data (*Güsewell et al.*, 2017), we decided to publish them as a whole so that they would be accessible to all scientists working with Swiss phenological data. The model comparison presented in Chapter 6 has meanwhile served to develop and implement an additional powerful automatic quality check at MeteoSwiss, but results may still contribute to the development of data verification procedures in other countries.

2 Data

2.1 The Swiss Phenology Network

The Swiss Phenology Network is a long-term citizen-based monitoring programme of MeteoSwiss initiated in 1951 (*Primault*, 1955; *Defila and Clot*, 2001). Data are the day of year at which individual plants or plant populations of a certain species reach a certain stage of their annual life cycle, such as leaf unfolding or full flowering. In this report, a phenological stage recorded on a particular plant species is called a phenophase. Phenophases are precisely defined by a combination of biological and statistical criteria, e.g. “50% of the leaves (on average two leaves per bud) are expanded to the point that the base of the petiole is visible” (*Brügger and Vassella*, 2003). The onset date of phenophases is assessed visually by volunteering observers based on regular observations of the plants (1–3 times/week, depending on the season). Observations are done at fixed locations spread across Switzerland, called ‘stations’. At each station, the same plants or plant populations are observed every year by one or a few persons. Each phenophase is recorded once at each station provided that the plant species is present at the location. The dataset analysed here includes 167 ‘active’ stations (stations currently monitored or for which a new observer is searched) and 69 phenophases. Information about each station is provided in Appendix 1, and information about each phenophase is provided in Appendix 2. In the following, the onset date of one phenophase at one station in a particular year is called ‘observation’. The sequence of observations for one phenophase at one station is called ‘time series’. Data analysis and the presentation of results in tables and figures will often combine all phenophases representing the same life cycle stage. The term ‘phenophase’ or ‘phase’ is used for species-specific results, and the term ‘stage’ or ‘phenological stage’ for results including all species.

The altitude of stations ranges from 200 to 1900 m a.s.l., with 24% of the stations located above 1000 m. Due to the topography of Switzerland, high-elevation stations are clustered in some parts of the country (Fig. 1a). Stations can also be classified by climatic regions (Fig. 1b), which are a geographic subdivision of Switzerland based on topography and associated with different seasonal patterns of temperature, precipitation and wind. There are five main regions (Jura, Plateau or central lowlands, northern Alps, central Alps and southern Switzerland, which are further subdivided into eastern, western and central parts (MeteoSwiss, unpublished data). This report only considers the five main regions to include a sufficient number of stations per region.

Stations were established between 1951 and 2010 (Fig. 2a). Approximately half of them provide uninterrupted time series, while some records are missing for the others (Appendix 1). Accordingly, stations differ widely in the number of years for which data are available (Fig. 2b). Between 8 and 69 phenophases are recorded at each station. Most stations monitor at least 50 phenophases (Fig. 2c).

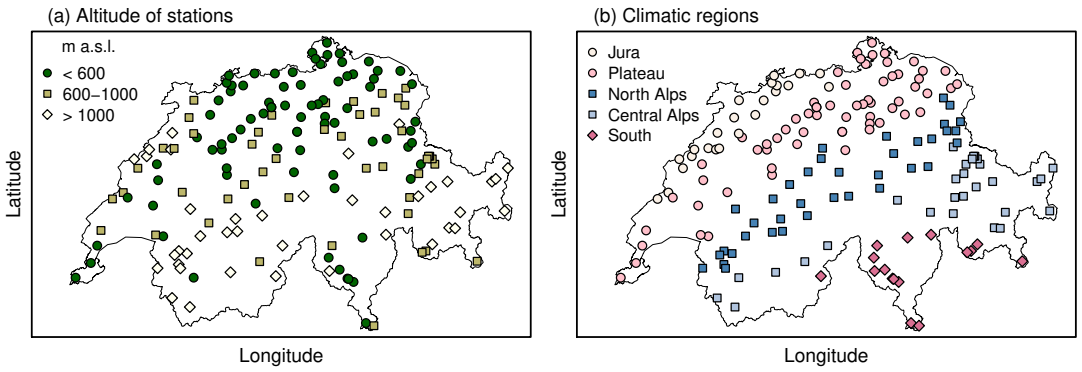


Figure 1: Location of the 'active' stations of the Swiss Phenology Network with their attribution to (a) altitudinal layers and (b) climatic regions.

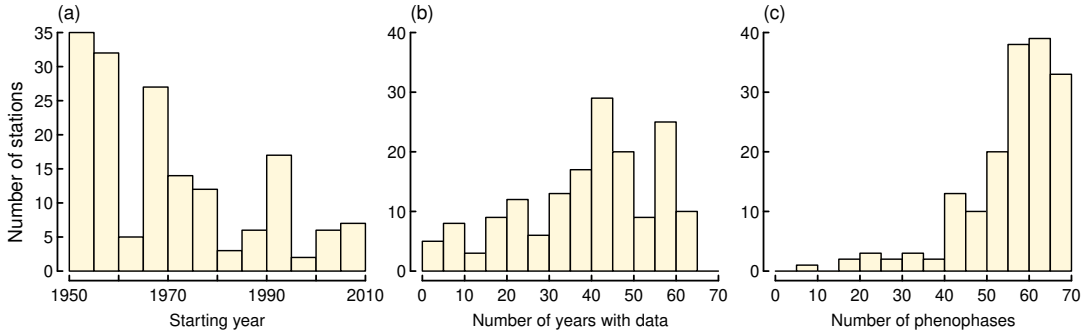


Figure 2: Frequency distribution of (a) the first year of data collection, (b) the number of years with data records, and (c) the number of phenophases recorded among the 167 'active' stations of the Swiss Phenology Network.

Most of the 69 phenophases are recorded at 120 or more stations (Appendix 2). Twelve phenophases recorded in fewer than 100 stations are excluded from this data analysis; they concern non-native or cultivated plant species restricted to the lowest elevations and southern Switzerland (*Castanea sativa, Robinia pseudoacacia, Vitis vinifera*). The remaining 57 phenophases encompass 23 plant species and 11 phenological stages. In this report, some stages are combined to obtain a simpler classification with only 7 stages and more species per stage: leaf unfolding or needle emergence (10 species), start of flowering (11), full flowering of trees and shrubs (11), full flowering of herbs (8), fruit maturity or harvest (4), leaf or needle colouring (8), and leaf or needle drop (5). Plant species are trees and shrubs, except for full flowering of herbs and hay harvest. In the following, the term 'full flowering' will refer to trees and shrubs, while 'flowering of herbs' or 'flowering' will be used for herbs. Furthermore, the terms 'leaf unfolding', 'leaf colouring' and 'leaf drop' will include needle emergence, needle colouring and needle drop, respectively.

Statistical analyses partly focus on four stages to simplify the presentation of results: leaf unfolding (including needle emergence), full flowering, flowering of herbs and leaf or needle colouring. The omission of three stages is only a small loss of information because start of flowering strongly correlates with full flowering (Fig. 3), fruit maturity correlates with full flowering while being recorded for few species, and leaf drop correlates with leaf colouring while being recorded for fewer species.

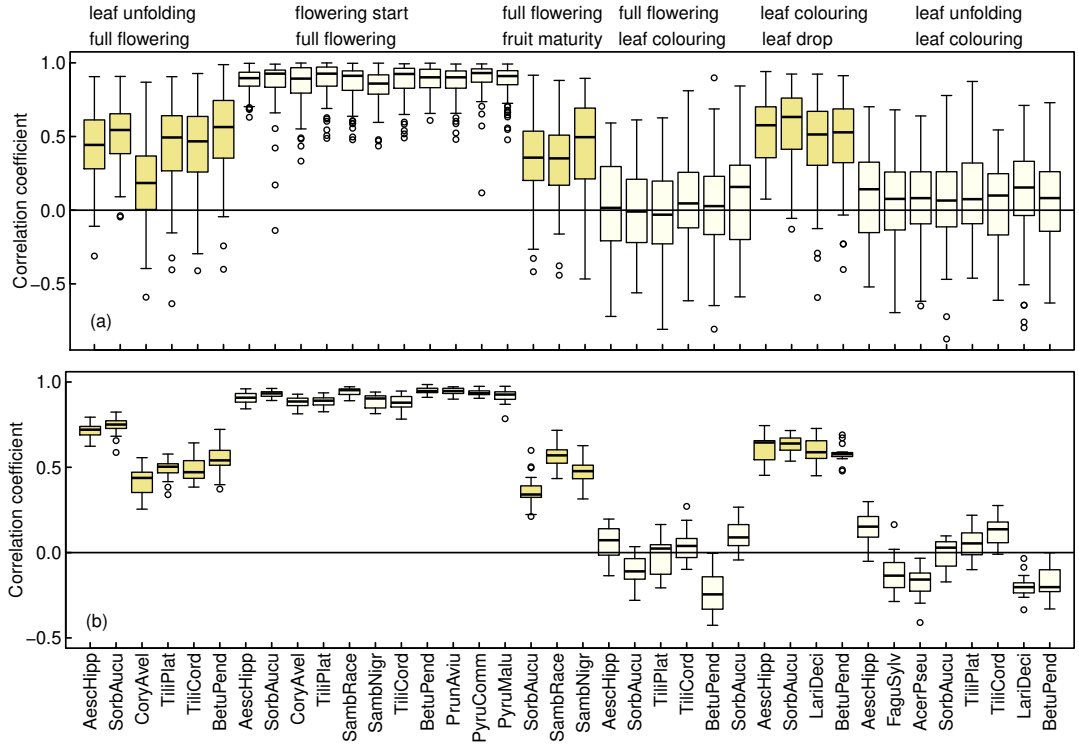


Figure 3: Spearman rank correlations between phenological stages of the same species. (a) Distribution of correlations across years, calculated for each station. Positive correlations indicate that the two stages tend be reached earlier or later than average in the same years, while negative correlations indicate that 'early years' for one stage correspond to 'late years' for the other stage. (b) Distribution of correlations across stations, calculated for each year. Positive correlations indicate that the two stages tend to be reached earlier or later than average at the same stations. Correlations between species for one phenological stage are represented in Appendix 3.

2.2 Data preprocessing

The original database extract from 30.09.2013 (`phaeno_active_stations.csv`) contains some duplicate entries (mostly from 2010), which have been removed. Furthermore, some records from the end of December (mostly in 2002) actually represent very early onset of Hazel flowering (*Corylus avellana*). These entries have been changed into negative values representing days up to 1 January of the following year. Similarly, instances of late leaf drop (in January or February) have been changed into days after 31 December.

Two subsets of data are considered in analyses: 'Recent' data include records from 1996–2012 for stations with at least 14 out of 17 years of data (138 stations). 'Long-term' data include records from 1970–2012 for 26 variables (those recorded from 1951) and for stations with at least 35 out of 43 years of data (108 stations). Most analyses are based on the recent data, which are the basis for future monitoring.

Plotting all time series for each phenophase reveals different types of outliers (see Fig. 4):

1. Some time series generally exhibit extreme values compared to all other stations. These are often short time series (only a few years), suggesting that the records were found to be incorrect after some time (misidentified plant species or untypical habitat).

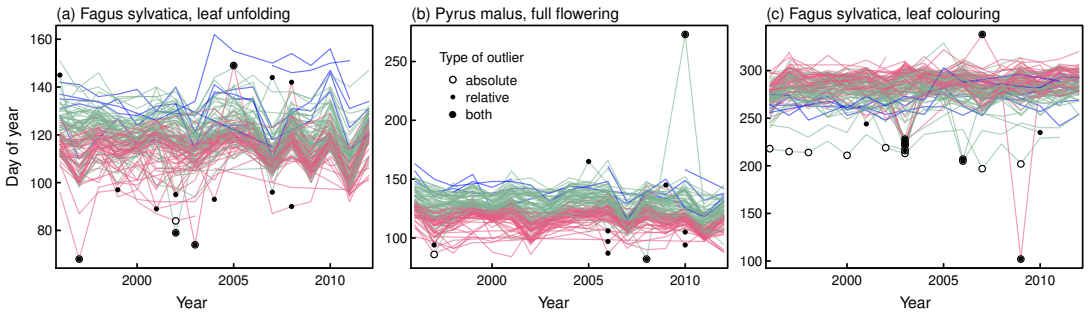


Figure 4: Time series with outliers for three selected phenophases. Line colours represent the three groups of stations formed by cluster analysis, based on which 'absolute' outliers were defined (see text). 'Relative' outliers are observations with large standardized residuals in linear models with additive effects of years and stations.

2. Some time series include individual extreme observations, probably because of an error during observations or data record, or alternatively, due to disease of the plant or other special events.

3. Finally, time series may include observations that are not extreme *per se*, but that deviate strongly from expectations based on the average phenology of the station or overall inter-annual variation. For example, a low-elevation station, usually with early phenology, may be as late as the uppermost stations in a certain year. Alternatively or in addition, a station may show a relatively late phenology in a year when phenology is generally early. Such outliers will be called 'relative' outliers.

To identify these different types of outliers in a simple and automatic way, two approaches are combined. The procedure is run separately for each phenophase and for the recent and long-term data, respectively.

To identify 'absolute' outliers (types 1 and 2 above) and outliers relative to a stations' average phenology, stations are grouped by cluster analysis: Based on the median value of each phenophase at each station, a Gower distance matrix[1] between stations is computed (function `daisy` in library `cluster`, *Mächler et al.*, 2013), and hierarchical agglomerative clustering is performed with complete linkage (function `hclust`). The resulting dendrogram reveals three groups of stations with contrasting altitude (similar to, but not identical with the altitudinal layers in Fig. 1). For each of the three groups of stations, means and standard deviations are calculated for each phenophase, and values outside the interval mean $\pm$ 3.5 sd are excluded from further data analysis.

To identify outliers relative to overall inter-annual variation (type 3 above), linear models with the additive, random effects of station and year are fitted to onset dates of each phenophase. The standardized residuals of these models indicate how much an observation deviates from what would be expected based on the additive effects of station and year. Observations with standardized residuals greater than 3.5 are excluded from further data analysis. In total, 723 out of 91358 observations (0.79%) are excluded from the recent data, and 610 out of 84614 observations (0.72%) from the long-term data. The fraction of data excluded is consistent with preliminary analyses of residuals from various models

[1] Gower's dissimilarity coefficient between two stations s and s' (based on p variables Y_k) is defined as $d_{ss'} = \frac{\sum_{k=1}^{p} \delta_{ss'k} d_{ss'k}}{\sum_{k=1}^{p} \delta_{ss'k}}$ i.e. the weighted mean of distances $d_{ss'k}$ with weights $\delta_{ss'k}$ over the p variables. The k-th variable's contribution to the distance, $d_{ss'k}$, is the absolute difference of y_{sk} and $y_{s'k}$, divided by the total range of variable Y_k. The weights $\delta_{ss'k}$ are zero if y_{sk} and/or $y_{s'k}$ is missing, and 1 otherwise. Thus, Gower distances can be compared even if the number and identity of available variables differs for each pair of stations.

Table 1: Descriptive statistics for recent data (1996–2012). The percentage of data available and mean dates were first calculated for each time series (one phenophase at one station) and then averaged over stations and species for each phenological stage. Standard deviations (sd total) were calculated across stations and years for each phenophase and then averaged for each stage. Variance components for the additive effects of stations, years and residual variation were obtained from random-effects models for each phenophase, converted into standard deviations, and averaged for each stage. All standard deviations are given in days.

Stage	% data	mean doy	sd total	sd stations	sd years	sd residual
Leaf unfolding	82	116	12.8	10.1	4.8	6.4
Flowering start	74	123	15.4	12.6	6.2	7.6
Full flowering	76	131	15.5	12.9	6.1	7.4
Flowering of herbs	77	141	17.7	14.7	5.4	9.3
Fruit maturity	73	206	21.6	17.5	5.0	12.5
Leaf colouring	76	283	13.7	9.2	3.6	9.6
Leaf drop	78	305	12.9	8.5	3.6	9.1

using half-normal plots, showing that roughly 1% of the residuals are more extreme than with a normal distribution. Fig. 4 shows examples of time series for three phenophases with the grouping of stations (coloured lines) and with observations excluded as outliers according to either criterion. Outliers are distributed rather homogeneously among phenophases (see Fig. 4), while the distribution among stations appears random (similar to Poisson).

After outlier exclusion, 90635 observations remain in the recent data, which represents 67.8% of all theoretically possible observations (138 stations, 57 phenophases, 17 years). In the long-term data, 84004 observations remain, i.e. 69.6% of all theoretically possible observations (108 stations, 26 phenophases, 43 years). Missing observations are due to missing time series (phenophases not recorded at a station) and missing or excluded data within existing time series. In the recent data, 6917 of the 7866 possible time series (87.9%) are represented by at least one observation, and in the long-term data 2563 of the possible 2808 time series (91.3%). Within these time series, data are available on average for 77.7% (76.2%) of the years in the recent and long-term data, respectively. Availability of data is slightly better for leaf unfolding than for the other stages (Table 1).

Table 1 provides descriptive statistics for the recent data (1996–2012), after exclusion of outliers. All statistics were calculated for individual phenophases; means per stage are given in Table 1. The standard deviation of individual phenophases across stations and years ranges from 9.7 to 26.5 days. This overall phenological variability can be decomposed into spatial, temporal and residual components using again linear models with the additive, random effects of station and year. Standard deviations among stations (7.6 to 23.3 days for individual phenophases) are about two times larger than those among years (2.9 to 13.4 days for individual phenophases). Residual variation (deviations from the additive effects of station and year) ranges from 4.9 to 15.1 days. Table 1 shows that variation among stations is most pronounced for flowering and fruiting, variation among years is smallest for the autumn phases, and residual variation is smallest for the spring phases. Results for the long-term data (1970–2012, not shown) are similar to those for the recent data, except that variation (sd) among years is 1–2 days larger.

3 Phenological variation in relation to altitude, climatic regions and spatial distance

3.1 Aims and methods

This chapter analyses how phenological variation depends on the stations' altitude, how it differs among climatic regions, and how it is related to the spatial distance between stations. The aim is to evaluate the relevance of these factors for the representativeness of the Swiss Phenology Network. For example, if differences among stations mainly depend on altitude, the Network should aim at a good representation of each altitudinal layer. If differences mainly depends on spatial distance, an even spatial distribution (avoiding clumps of nearby stations) is particularly important.

Data from each station are time series of onset dates for different phenophases, hence the temporal dimension must be included in the comparison of stations. This will be done at three levels, which represent different perspectives on phenological variation through time: First, the temporal dimension is excluded by analysing mean onset dates over time. Second, phenological variation from year to year is considered, and third, long-term trends. At each level, a summary statistic is computed for each phenophase at each station, which is then related to altitude, climatic regions and spatial distance. The analyses are described in detail only for mean onset dates as they are the same for the two other levels. Some specific analyses are additionally performed at each level to better describe and illustrate patterns of phenological variation. Most analyses are performed separately for each phenophase. Results are partly summarized by phenological stage in the report, while detailed results for each phenophase are presented in the Appendix. Some results are only presented for the four focal stages, i.e. leaf unfolding, full flowering, flowering of herbs and leaf colouring.

Mean onset dates

Based on the recent data (1996–2012), mean onset dates of each phenophase at each station are computed provided that data from $\geq$ 6 years are available. Preliminary analyses with the long-term data or with data from individual years yielded similar results, but with lower precision, and are not shown here. Relationships of mean onset dates with altitude are analysed with linear regression. The residuals of regression models are used to identify stations with particularly early or late phenology relative to their altitude.

Differences among climatic regions are analysed with linear models including the effects of altitude (continuous), region (categorical with 5 levels) and their interaction. Models must include altitude

because there is substantial covariation between altitude and regions in Switzerland: the mean altitude of stations is 540.5 m a.s.l. in the Plateau but 1131.4 m in the Central Alps. Overall, 34.9% of the variation in stations' altitudes is among regions. Hence, the effects of altitude and region on plant phenology are partly confounded. To describe their relative influence, the total sum of squares of the onset dates of each phenophase is decomposed additively into fractions explained by local variation in altitude (effect of altitude alone), regional variation in altitude (joint effect of altitude and region), regional differences unrelated to altitude (effect of region alone), and region-specific effects of altitude (interaction of altitude and region). This decomposition is obtained by calculating sequential sums of squares from models starting either with the effect of altitude (fit 1: $Y \sim$ altitude $*$ region) or with the effect of regions (fit 2: $Y \sim$ region $*$ altitude). From these models we obtain the following effects:

- local variation in altitude (altitude alone)	SS_{altitude} (fit 2)
- regional variation in altitude (joint)	SS_{altitude} (fit 1)$-SS_{\text{altitude}}$ (fit 2)
- regional differences unrelated to altitude (region alone)	SS_{region} (fit 1)
- regional differences in the response to altitude (interaction)	$SS_{\text{altitude:region}}$ (fit 1)

The SS fractions are expressed as percentages of the total sum of squares of each phenophase. Means and sd of these percentages are calculated for each of the four focal stages. Tests of significance for the effects are obtained from the sequential ANOVA tables. No test of significance is obtained for the effect of regional variation in altitude because this is not a model effect and does not appear in either ANOVA table.

Linear models with the effects of altitude and region without interaction are used to derive predicted means $\pm$ sd for each phenophase in each region at the overall mean altitude of 781.7 m a.s.l., as a concrete measure of the regional differences unrelated to altitude.

Spatial dependence is analysed in two ways. First, Mantel *r*, a measure of spatial autocorrelation, is computed for each phenophase. This is the correlation between spatial distances and differences in phenology for all pairs of stations, which can be caused by large-scale trends or local spatial dependence. Mantel *r* is also computed from the residuals of regressions against altitude to see whether spatial correlation of phenology is due to the spatial dependence of altitude in Switzerland or other spatial processes.

Second, to better describe the spatial processes, spatial surfaces are fitted using a kriging model with the effects of altitude, a large-scale spatial trend and spatial correlation: $y_s = \beta z_s + p(x_s) + f(x_s) + \epsilon_s$, where y_s is the mean onset date of a phenophase at station s, z_s is the station's altitude, β a regression coefficient, $p(.)$ a polynomial function (degree 2), x_s the stations's spatial coordinates, $f(.)$ a Gaussian field with mean = 0 and covariance function k, and ϵ_s an i.i.d. Gaussian error with variance σ^2. An exponential covariance structure is assumed for $f(.)$, i.e. the covariance of two observations at distance h is $k(h) = \rho \cdot \exp(-h/\theta)$, where ρ is a scale parameter ('sill'), and θ is a range parameter. Altitudes are centered, so that the spatial surface is fitted at the overall mean altitude. Contour maps of the spatial surfaces visualize spatial patterns. The range and the importance of spatial correlation are characterized by θ and by the smoothing parameter $\lambda = \sigma^2/\rho$, respectively.

Models are fitted with the function `Krig` in the R package `fields` (*Nychka et al.*, 2013). This function does not estimate the range parameter θ and requires it to be specified. To choose a suitable value, a

sequence of models is fitted with θ values ranging from 15 km (1% quantile of the observed pairwise distances among stations) to 270 km (99% quantile of the distances). The value leading to the smallest estimate of residual error (σ) is chosen.

Interannual variation

Interannual variation describes how phenological onset dates in individual years deviate from the long-term mean for a certain phenophase and station. Positive deviations correspond to years with relatively late phenology, while negative deviations represent years with relatively early phenology. The average size of these deviations (standard deviation) indicates the variability of onset dates over time. Based on data from 1996–2012 (recent data), standard deviations are computed for each phenophase and station provided that data from $\geq$6 years are available. The variability (standard deviation) of each phenophase is then related to the altitude, climatic regions and spatial distance of stations as described above for means.

In addition, interannual variation is analysed using mixed models to see whether interannual patterns (the sequence of early and late years) differ in relation to altitude or climatic regions. The basic model for the onset date of a certain phenophase (y) in year i at station s is a mixed model with the fixed effect of year (J_i) and the random effect of station (α_s): $y_{is} = J_i + \alpha_s + \epsilon_{is}$.[2] Years are treated as a categorical variable (trends are analysed in the next section). For a joint analysis of all phenophases belonging to one stage, the model additionally includes the fixed effect of species (S_j) and a species-specific effect of year:

$$y_{ijs} = S_j + J_{ij} + \alpha_s + \epsilon_{ijs} \quad (1)$$

There is no intercept, and the effect of years is nested within species and coded with sum-to-zero contrasts, so that coefficients represent deviations from the mean onset date for each species.

Relationships with altitude are analysed by comparing three altitudinal layers ($<$600 m, 600–1000 m, $>$1000 m, see Fig. 1) to facilitate the interpretation and graphical representation of results. Interannual patterns are compared among altitudinal layers and among climatic regions by including either factor in the basic model as fixed effects, similar to model (1). The model now describes an altitude-specific effect of year or a region-specific effect of year, respectively. For the comparison of altitudinal layers, the model is: $y_{i\ell s} = L_\ell + J_{i\ell} + \alpha_s + \epsilon_{i\ell s}$, where L_ℓ is the effect of altitudinal layer ℓ. Climatic regions are compared with an analogous model: $y_{irs} = R_r + J_{ir} + \alpha_s + \epsilon_{irs}$, where R_r is the effect of region r.

Spatial and temporal dependence is analysed by allowing errors to be correlated in model (1). Spatial correlation is modelled within groups defined by years and plant species, while temporal correlation is modelled within groups defined by stations and plant species; errors from different groups are assumed to be independent. Spatial correlation can arise from small-scale variation in climatic factors or other ecological factors that affect neighbouring stations similarly. Temporal correlation can result from changes in the observers, plant individuals, plant age or soil conditions, i.e. factors whose strictly local influence can change over periods of several years. For ease of computation and interpretation, spatial and temporal correlations are analysed separately. Models are fitted through generalized least squares estimation (function `gls` in package `nlme`; see Chapter 5.3.2 in *Pinheiro and Bates*, 2000). Because

[2] Capital letters denote fixed effects and Greek letters denote random effects.

random effects are not allowed, stations are exceptionally treated as fixed effects. A mixed model with correlation structure (function `lme`) cannot be used here because it does not allow the grouping structure to differ from the random effects.

For spatial correlation, examination of sample semivariograms and model comparison with AIC leads to the choice of an exponential correlation function with nugget effect. Thus, the correlation of within-group (ij) errors from two stations (s and s') at horizontal distance h ($h > 0$) is modelled as cor$(\epsilon_{ijs}, \epsilon_{ijs'}) = f(h, c_0, \theta) = (1 - c_0)$ exp$(-h/\theta)$, where $1 - c_0$ is the correlation of immediately neighbouring observations ($h \downarrow 0$), which is less than 1 due to the nugget effect, and θ is the spatial range. The effective range 3θ is the distance at which the correlation has dropped to less than 5% of its maximal value. Temporal correlation is modelled as an autoregressive process of order 1 (AR1). The importance of spatial and temporal correlation is judged from the difference in AIC between models with and without correlation structure, the difference in standard errors of estimated coefficients between the two models, and parameters of the correlation functions.

Long-term trends

Phenological trends are analysed with the long-term data, i.e. data from 1970–2012 for 26 phenophases. The trend is estimated for each station and phenophase through simple linear regression of onset dates against years provided that $\geq$10 years of data are available. The estimated trend is the regression slope, expressed in days per decade. A few extreme outliers representing implausible trends (delays of more than 10 days/decade or advances of more than 20 days/decade) are excluded from further analysis. Trends are related to altitude, climatic regions and spatial distance of stations as described above for means.

Average trendlines for each phenophase (based on all stations) are obtained from mixed models including stations as random effects. Confidence intervals for these trendlines and standard errors of estimated trends are derived from the variance-covariance matrix of fixed effects. This approach ignores the variance due to random effects, but parametric bootstrapping for a few phenophases showed this additional variance to be negligible due to the large number of stations included (mostly $>$100).

Note on *p*-values and 'significance'

Several tables and figures contain reports of *p*-values or indications about 'significant' effects ($p < 0.05$). This is done in an explorative sense to provide a simple and easily comparable indicator for the presence and size of effects. These tests do therefore not establish significance in the classical sense of hypothesis testing. For the same reason, no correction for multiple testing is applied even if tests are carried out separately for multiple phenophases.

Table 2: Mean phenological onset dates in relation to altitude and spatial distance: correlation (Pearson's *r*) between mean onset dates 1996–2012 and the altitude of stations, slope (days/100 m) and residual standard error (days) of linear regression, spatial correlation (Mantel *r*) of mean onset dates, and spatial correlation of regression residuals (autocorrelation left after removing the altitudinal trend). Correlations were calculated for each phenophase. Means (and sd) per stage are given in the table.

Stage	relationship with altitude			spatial correlation	
	correlation	slope	error	Mantel *r*	resid. Mantel *r*
Leaf unfolding	0.79 (0.04)	2.50 (0.30)	6.17 (1.02)	0.10 (0.03)	0.02 (0.03)
Flowering start	0.73 (0.12)	2.95 (0.37)	8.27 (2.83)	0.15 (0.04)	0.09 (0.07)
Full flowering	0.76 (0.11)	3.06 (0.30)	7.93 (2.92)	0.15 (0.04)	0.08 (0.08)
Flowering of herbs	0.63 (0.26)	2.64 (1.11)	10.66 (5.09)	0.09 (0.07)	0.03 (0.04)
Fruit maturity	0.58 (0.19)	2.95 (0.95)	14.01 (4.00)	0.09 (0.06)	0.05 (0.05)
Leaf colouring	–0.21 (0.19)	–0.57 (0.56)	9.27 (1.11)	0.01 (0.05)	–0.01 (0.03)
Leaf drop	–0.23 (0.20)	–0.60 (0.52)	8.44 (0.73)	0.06 (0.04)	0.04 (0.03)

3.2 Results

Mean onset dates

Mean onset dates of most phenophases are linearly related to altitude (Appendix 4, Fig. 5), but the slope and strength of the relationship differ among phenological stages. Leaf unfolding and full flowering are strongly positively related to altitude, with an average delay of 2.5 to 3.1 days per 100 m higher elevation (Table 2, Fig. 5a, b). Flowering of herbs and fruit maturity have a similar altitudinal trend but slightly weaker correlations (Table 2). Leaf colouring and leaf drop are weakly negatively related to altitude, with 0.58 days earlier dates per 100 m higher elevation (Table 2, Fig. 5c).

In models combining the effects of altitude and climatic regions, altitudinal variation (both local and regional) accounts for 51–69% of the variation in mean onset dates of leaf unfolding, 28–79% of the variation in tree flowering dates, and a smaller fraction of variation for the other stages (Table 3). For most phenophases (except for leaf colouring and leaf drop) the effect of local variation in altitude is statistically significant ($p < 0.05$) and explains 2–4 times more variation than the effect of regional variation in altitude. Differences among regions that are unrelated to altitude explain only 0–12% (mean: 4.6%) of the variation in onset dates but are still statistically significant for most spring phenophases. The interaction between the effects of altitude and climatic regions also explains only 0–12% (mean: 3.0%) of the variation and is significant for a few phenophases.

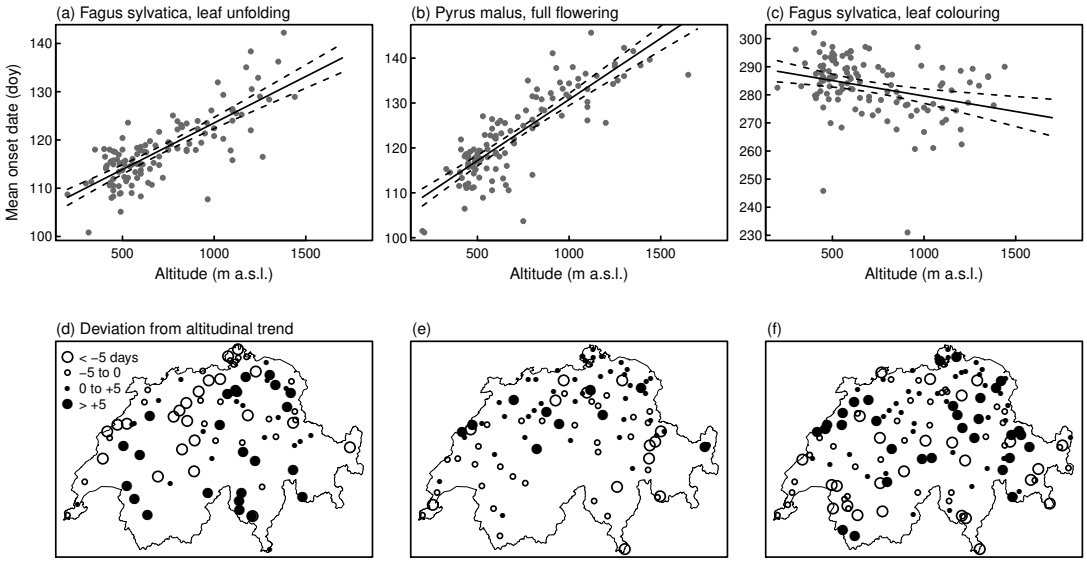


Figure 5: Regression of mean onset dates (1996–2012) against altitude for three selected phenophases (a–c) and spatial distribution of deviations from the overall altitudinal trend, i.e. deviations from the regression line (d–f).

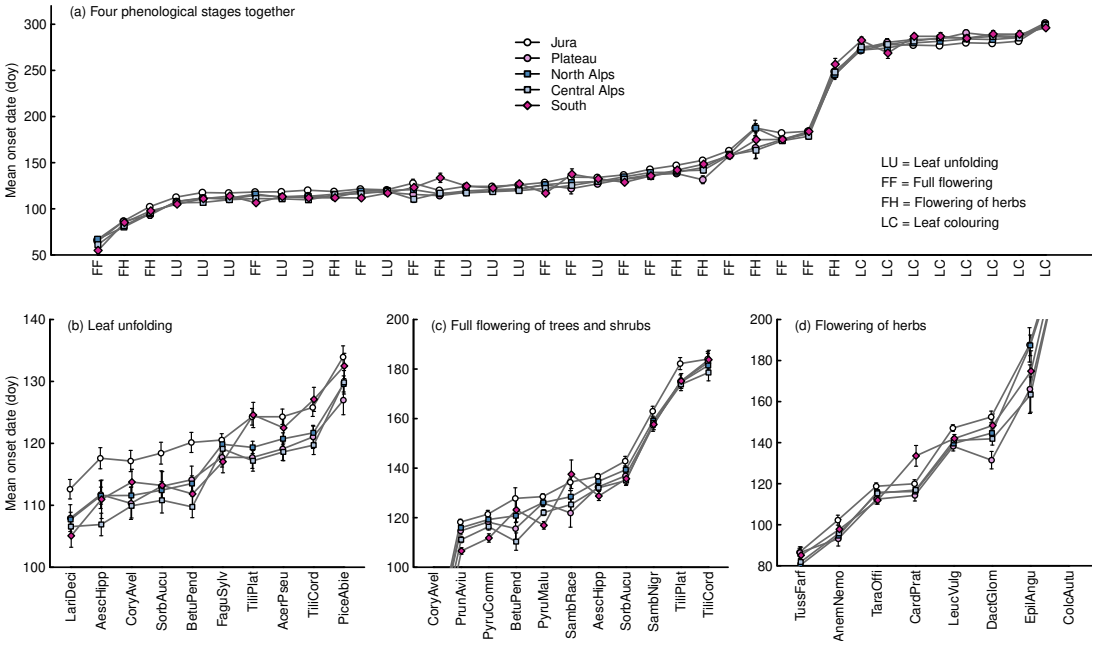


Figure 6: Predicted mean onset date (1996–2012) of each phenophase in each region at the overall mean altitude of 781.7 m a.s.l. Predictions are derived from linear models with the effects of altitude and region. Error bars (if visible) show ± 1 se. Panel (a) combines data from the four focal stages in chronological order, while panels (b)–(d) show close-ups for three stages with greater resolution on the y axis (note the different scales). The order of phenophases within stages in panel (a) is identical to that in panels (b)–(d).

Table 3: Effects of altitude and regional differences on mean phenological onset dates (1996–2012). The total variation (sum of squares) in each phenophase is decomposed into fractions explained by altitude, the five climatic regions, their interaction and residual variation. Altitude varies both within region ('local') and among regions ('regional'); the corresponding fractions of variation are given separately. The fraction of variation explained purely by regional differences is obtained after accounting for altitudinal differences. All fractions are expressed as percentages of total variation. Percentages were calculated for each phenophase, and means (with sd) per stage are given in the table; *n* is the number of phenophases (i.e. species) per stage. Superscript numbers indicate the number of phenophases for which an effect is significant (ANOVA, $p < 0.05$); there is no test of significance for the effect of regional variation in altitude.

Stage	*n*	altitude		region	interaction	residual
		local	regional			
Leaf unfolding	10	44.2 (5.3)^{10}	18.5 (5.1)	4.4 (2.3)^{8}	1.2 (1.5)^{1}	31.7 (4.6)
Flowering start	11	43.0 (12.7)^{11}	11.8 (7.2)	5.1 (4.0)^{7}	2.0 (2.9)^{1}	38.0 (16.0)
Full flowering	11	45.6 (12.7)^{11}	13.4 (6.6)	5.4 (3.2)^{7}	1.8 (1.7)^{0}	33.8 (16.1)
Flowering of herbs	8	31.5 (18.4)^{7}	14.7 (8.6)	3.7 (1.6)^{2}	3.6 (2.7)^{4}	46.5 (24.3)
Fruit maturity	4	24.3 (17.5)^{4}	12.2 (5.8)	3.4 (2.3)^{1}	4.7 (3.5)^{2}	55.4 (24.5)
Leaf colouring	8	4.2 (3.7)^{4}	3.4 (4.2)	5.1 (2.5)^{1}	4.0 (2.1)^{1}	83.3 (8.4)
Leaf drop	5	5.9 (6.1)^{3}	2.7 (3.8)	3.9 (1.8)^{0}	5.0 (3.4)^{0}	82.4 (7.0)

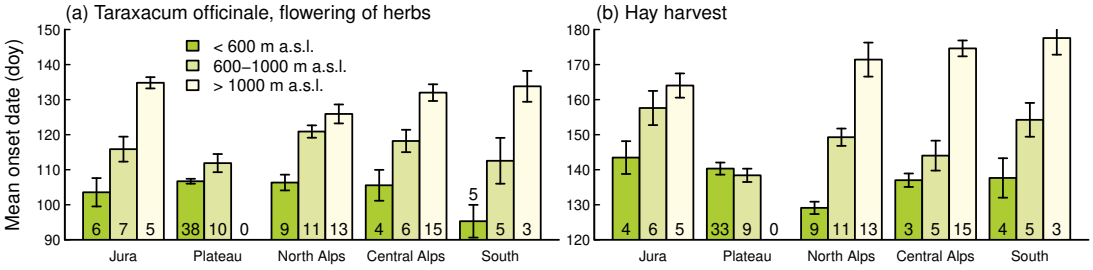


Figure 7: Differences in mean onset dates among altitudinal layers and climatic regions for two phenophases with significant interaction effects (Table 3). Mean onset dates (1996–2012) were calculated for each station; bars show means $\pm$ se per altitudinal layer and climatic region. The number of stations is given in each bar.

Differences among regions are illustrated in Fig. 6 by predicting the mean onset date of each phenophase in each region at the overall mean altitude of 781.7 m a.s.l. The Jura tends to be late, while the central Alps or southern Switzerland tend to be early, but some phenophases deviate from this pattern. For most variables, regions differ by less than 10 days from each other (Fig. 6), while the effect of altitude can exceed 40 days (Fig. 5b). The combined effect of altitude and climatic regions is illustrated in Fig. 7 for two phenophases with a significant interaction effect. For these phenophases, differences between altitudinal layers depend on the region (e.g. the two lower layers differ less in the Plateau than in the other regions), and differences between regions partly depend on the altitude (e.g. flowering of *Taraxacum officinale* occurs particularly early in southern Switzerland at low altitude but not at high altitude, see Fig. 7a).

The spatial distribution of mean onset dates across Switzerland primarily reflects their dependence on altitude. Because altitude is spatially correlated across Switzerland (Mantel r = 0.19), phenophases that correlate with altitude also present some spatial dependence (Table 2), while deviations from

Table 4: Stations with particularly early or late phenology when considering all phenophases or the four stages, based on linear regressions of mean onset dates (1996–2012) of each phenophase against the altitude of stations. Stations are classified as early (late) phenology if their mean onset date is more than 5 days earlier (later) than predicted at their altitude for at least 75% of the phenophases. See Appendix 1 for information about stations and Appendix 5 for the distribution of deviations from altitudinal trends.

Stage	late stations	early stations
All	1722	8043, 9858
Leaf unfolding	201, 521, 1668, 1722, 1761, 3075, 3761, 5231, 6299, 6392, 8539, 9778	588, 639, 642, 1941, 4121, 5351, 5495, 6905, 7957, 8043, 9858, 9981
Full flowering	1722, 3761, 5051, 5231, 6352, 7877, 9851	588, 639, 876, 1941, 7069, 8043, 9402, 9403, 9981
Flowering of herbs	1239, 1444, 1722, 1761, 3761, 6238, 6299, 7715	56, 2855, 4121, 7573, 7801, 7877, 8043, 9402, 9449, 9709, 9858, 9932
Leaf colouring	56, 338, 642, 876, 1011, 1892, 2201, 2778, 3629, 4589, 5289, 5529, 5742, 5871, 6173, 6371, 6539, 7330, 7642, 7964, 8029, 9353, 9402	2018, 3799, 5051, 5231, 5469, 5495, 6069, 6299, 6326, 6392, 6469, 6592, 6765, 6905, 6993, 9709, 9858

the altitudinal trend (residuals of linear regression models) mostly present little spatial dependence (Table 2). In the following, we focus on the spatial distribution of deviations from the altitudinal trend. For some phenophases, positive and negative deviations are distributed rather homogeneously over the country, e.g. leaf unfolding of *Fagus sylvatica* in Fig. 5d. In contrast, full flowering of *Pyrus malus* has mostly negative deviations (early flowering) in southern Switzerland and mostly positive deviations in northern Switzerland (Fig. 5e). This spatial pattern is reflected by a relatively strong correlation of residuals (Mantel $r = 0.11$).

Deviations from altitudinal trends can be used to identify 'early' and 'late' stations relative to their altitude. However, these deviations are rarely consistent for a particular station. If all phenophases are considered, most stations present deviations ranging from negative to positive, and the distribution of these deviations is similar for most stations (Appendix 5). Only a single station has mostly large positive deviations (late phenology), and two stations have mostly large negative deviations (early phenology, Table 4). For individual stages, differences among stations are more pronounced (Appendix 4), and a number of stations with early or late phenology can be identified for each stage (Table 4). These stations are not concentrated in particular regions, as early and late stations can be found in all regions of Switzerland (Appendix 5).

Spatial models decompose the spatial field of deviations from the altitudinal trend into large-scale trends (polynomial response surface) and small-scale dependence (Gaussian field). The relative importance of these two processes appears to vary among phenophases. This is illustrated in Fig. 8 by plotting the spatial distribution of predicted onset dates at mean altitude (i.e. overall mean onset date + deviations from altitudinal trend) for four of the phenophases. In three cases the large-scale trend dominates, as indicated by simple patterns of contour lines and a large smoothing parameter. The predicted date of leaf unfolding in *Fagus sylvatica* varies little in space, which means that variation in onset dates for this phenophase is almost entirely determined by altitude and purely local (spatially unstructured) variation (Fig. 8a). The predicted date of full flowering in *Pyrus malus* shows a strong latitudinal gradient, which only partly corresponds to climatic regions (Fig. 8b). The predicted flowering date of *Anemone nemorosa* is earlier in the Plateau and northern Alps than in the Jura, central and southern Alps, i.e. in this case the large-scale trend is well described by the climatic regions (Fig. 8d). Unlike the

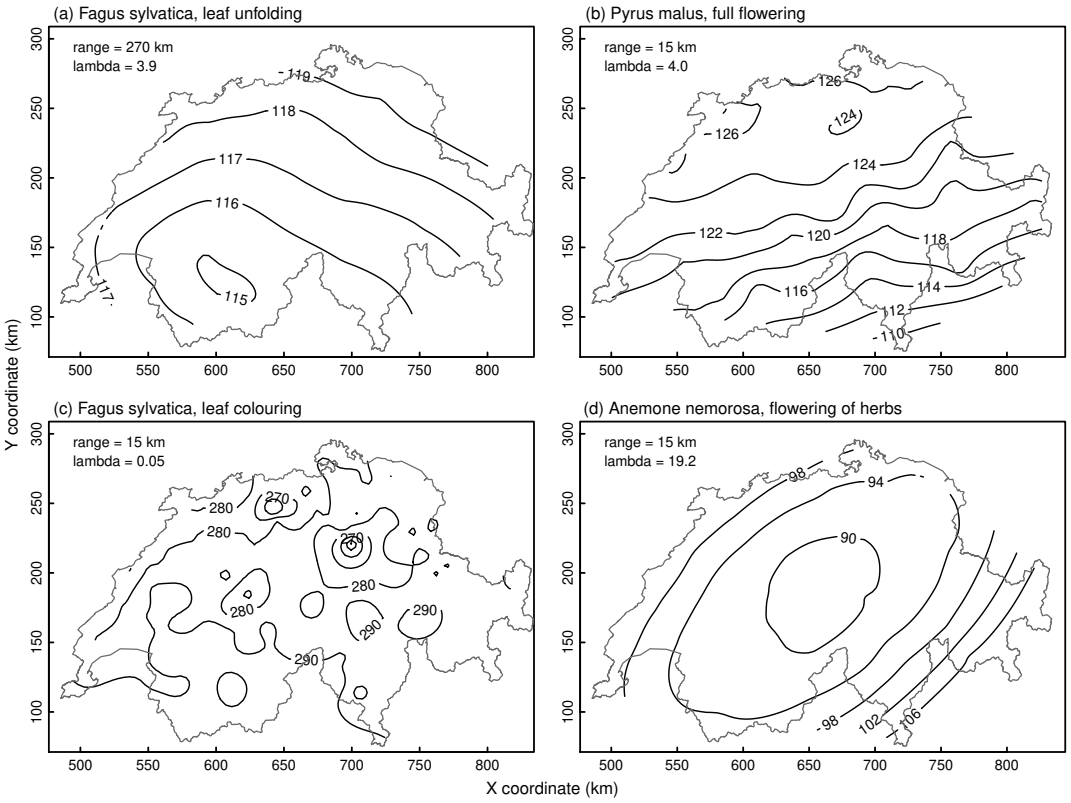


Figure 8: Spatial fields of predicted mean onset dates (1996–2012) at the overall mean altitude for four of the phenophases (a–d). Predictions are derived from kriging models with the effects of altitude, large-scale trend (polynomial response surface) and small-scale dependence (Gaussian field). Contour lines visualize the combined effects of large-scale trends and small-scale dependence. The range of spatial dependence and the smoothing parameter (residual variance relative to variance of the Gaussian field) are given in each graph.

other phenophases, the predicted date of leaf colouring in *Fagus sylvatica* shows no large-scale trend but some small-scale dependence (small smoothing parameter, Fig. 8c). However, the range of this spatial dependence (15 km) is so small that it includes only 1% of the pairs of stations; hence variation is purely local (spatially unstructured) for the majority of stations.

In conclusion, mean phenological onset dates in Switzerland primarily depend on altitude, followed by large-scale spatial trends, which may or may not correspond to the established classification into climatic regions. Small-scale spatial dependence plays only a minor role with the current density of stations, and a large part of the variation is purely local.

Interannual variation

Interannual variation is considerable, with a range of 15–20 days for most phenophases based on annual means of all stations (Fig. 9). Early and late years alternate in an irregular pattern. Dates of leaf unfolding show almost identical patterns in all species (Fig. 9a), whereas interannual patterns in flowering dates are more species-specific (Fig. 9b, c). The most deviant species are the very early flowering and variable tree *Corylus avellana* and the late flowering herbs *Epilobium angustifolium* and *Colchicum autumnale*.

At individual stations, interannual variability (standard deviation of onset dates from 1996–2012) is on average 9.6 days; this is similar for all phenological stages but differs slightly more among species (Table 5). The variability of a phenophase at individual stations typically ranges from 2 to 30 days (not shown), so that the standard deviation of variabilities (over stations) is on average 3.0 days (Table 5). Interannual variability is only weakly and inconsistently related to altitude or spatial distance (Table 5).

Altitude and climatic regions jointly explain on average 18% (range: 2–40%) of the variation among stations in interannual variability of individual phenophases (Table 6). Differences among climatic regions and region-specific effects of altitude (interaction effects) are most important. Fig. 10 represents differences in variability for two phenophases with relatively strong regional and interactive effects. In both phenophases, interannual variability tends to increase with altitude in the Jura, northern Alps and southern Alps, but the relationship with altitude is opposite in the central Alps. This pattern is found generally for leaf unfolding and full flowering and leads to a slight overall increase in variability with altitude (cf. Table 5). Table 6 suggests that variability differs among climatic regions, but this holds only for individual phenophases. Regions with higher or lower variability actually differ among phenophases, so that on average, regions do not differ in phenological variability.

Table 5: Phenological variability (standard deviation of years 1996–2012) in relation to altitude and spatial distance: mean variability and standard deviation of variabilities between stations, correlation with the altitude of stations (Pearson's *r*) and spatial correlation (Mantel *r*). Statistics were calculated for each phenophase. Means (and sd) of each phenological stage are given in the Table.

Stage	mean variability (days)	sd of variability (days)	correl. with altitude (*r*)	spatial correl. (Mantel *r*)
Leaf unfolding	8.94 (2.05)	2.78 (1.82)	0.15 (0.21)	0.05 (0.07)
Flowering start	9.29 (2.63)	2.74 (0.85)	0.02 (0.17)	0.05 (0.07)
Full flowering	9.68 (2.99)	2.92 (1.03)	0.07 (0.19)	0.07 (0.04)
Flowering of herbs	10.39 (1.13)	3.72 (0.88)	–0.11 (0.15)	0.06 (0.05)
Fruit maturity	9.54 (0.10)	3.47 (0.57)	0.03 (0.11)	0.03 (0.02)
Leaf colouring	9.60 (2.29)	2.98 (0.99)	0.00 (0.19)	0.04 (0.04)
Leaf drop	10.14 (2.84)	3.25 (1.84)	0.06 (0.17)	0.08 (0.04)

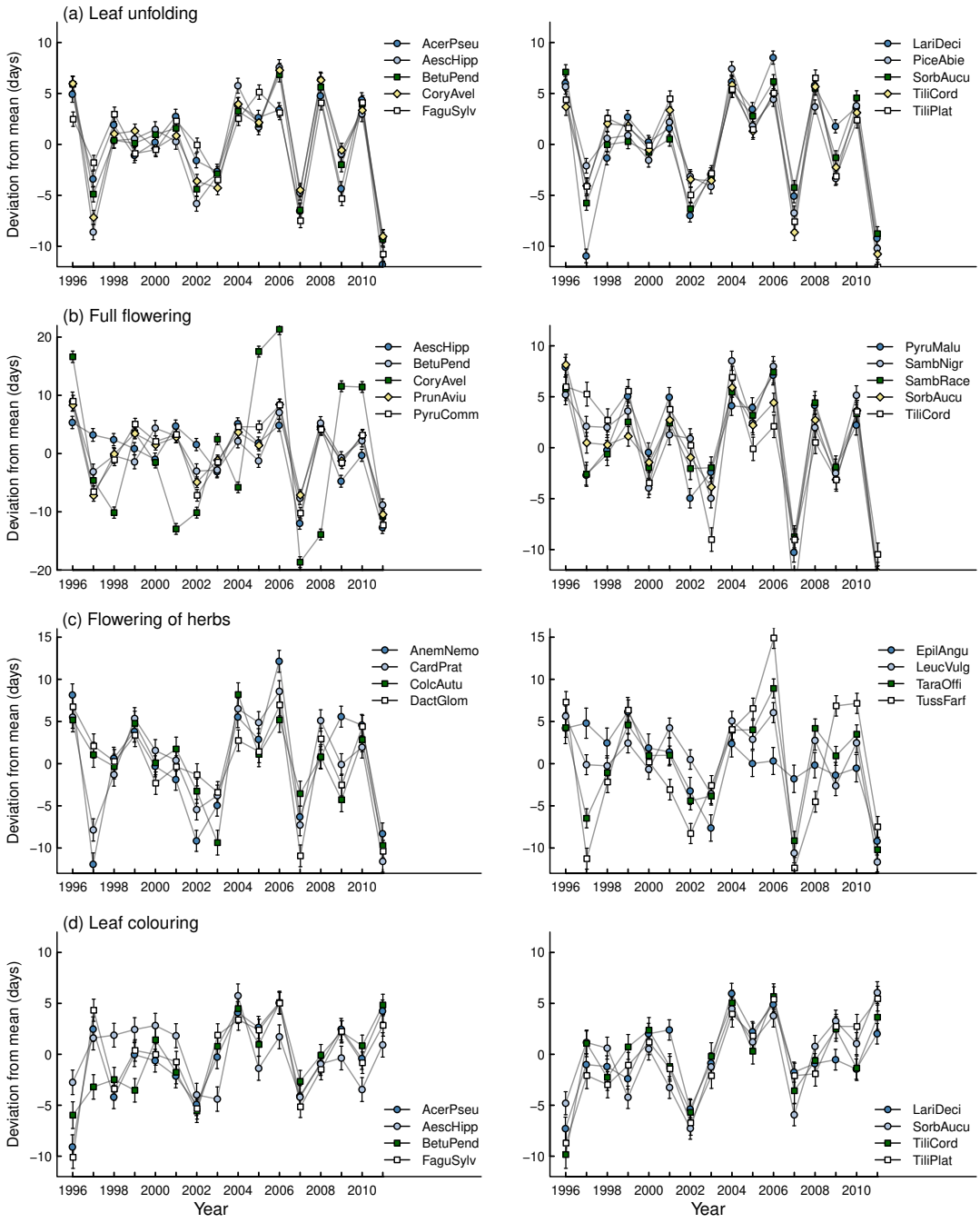


Figure 9: Annual deviations from mean onset dates for each species (means ± se over stations), derived from mixed models for the four focal phenological stages (a–d) with species and years as fixed effects and stations as random effects. Deviations in 2012 are the negative sum of deviations of the other years and therefore not represented. Graphs show that annual deviations of a phenological stage are generally similar in all species, with a few exceptions.

Table 6: Effects of altitude and regional differences on phenological variability (standard deviation of years 1996–2012), described by the percentage of variation explained by each factor and their interaction. Altitude varies both within region ('local') and among regions ('regional'). Means (and sd) of the phenophases belonging to each stage are given; *n* is the number of phenophases (i.e. species) per stage. Superscript numbers indicate the number of phenophases for which an effect is significant (ANOVA, $p < 0.05$). See Table 3 for further details.

Stage	*n*	altitude		region	interaction	residual
		local	regional			
Leaf unfolding	10	1.3 (1.5)^4	2.1 (2.5)	9.5 (6.6)^4	8.5 (9.3)^2	78.6 (8.5)
Flowering start	11	2.3 (1.7)^3	–0.2 (2.2)	8.4 (4.3)^7	6.5 (4.2)^3	83.0 (5.7)
Full flowering	11	3.6 (3.3)^4	0.9 (2.7)	8.2 (5.7)^4	6.1 (3.4)^3	81.2 (8.8)
Flowering of herbs	8	2.4 (2.9)^0	–0.7 (1.4)	5.7 (4.2)^3	2.9 (1.5)^2	89.7 (6.0)
Fruit maturity	4	2.0 (1.6)^0	–0.7 (1.2)	11.3 (4.1)^2	5.4 (5.8)^0	81.9 (3.1)
Leaf colouring	8	6.1 (5.4)^4	2.2 (2.9)	9.4 (3.7)^2	5.6 (3.0)^2	76.7 (5.8)
Leaf drop	5	2.3 (1.8)^3	–0.9 (1.8)	9.9 (2.8)^3	5.9 (2.1)^0	82.8 (2.9)

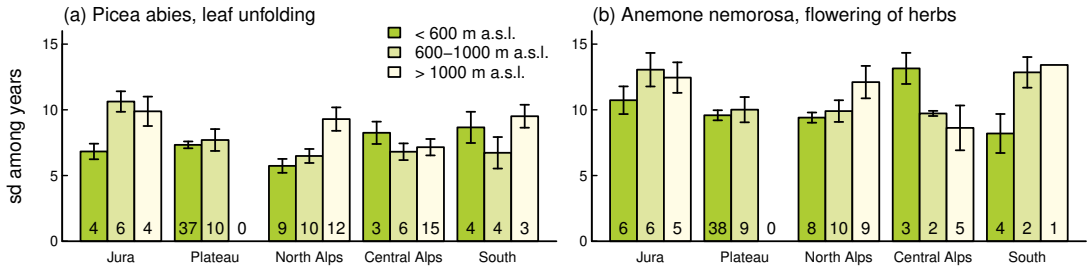


Figure 10: Differences in phenological variability among altitudinal layers and climatic regions for two selected phenophases. Variability (standard deviation of years 1996–2012) was calculated for each station; bars show means $\pm$ se per altitudinal layer and climatic region; the number of stations is given in each bar.

While interannual variability (over all years) does not differ consistently among altitudinal layers or climatic regions, deviations in individual years can differ substantially. Such altitude- or region-specific deviations are illustrated for three selected phenophases in Fig. 11. If we consider individual years with particularly early or late onset dates, we often find one altitudinal layer or one climatic region to be more 'extreme' than the others. However, the identity of these 'extreme' layers or regions changes from year to year, so that no altitudinal layer or climatic region consistently behaves in a specific way. We may note that southern Switzerland often deviates from the other regions but again, not in a consistent way.

Interannual variability presents only weak spatial correlation (Table 5) but deviations of individual phenophases in individual years are spatially correlated. This is seen by including a spatial correlation structure in models describing interannual deviations for each of the focal stages. AIC values are reduced compared to models with uncorrelated residuals, i.e. the predictive power of models is improved (Table 7). Spatial correlation increases the variance of estimated coefficients for annual deviations: standard errors of coefficients estimated from a model that accounts for spatial correlation are up to three times larger that those obtained from a model that ignores spatial correlation (Table 7). For leaf colouring, the effect of spatial correlation is small, which is related to a small effective range of the correlations (16.8 km): Only 2% of the distances between stations are within this range, so that most

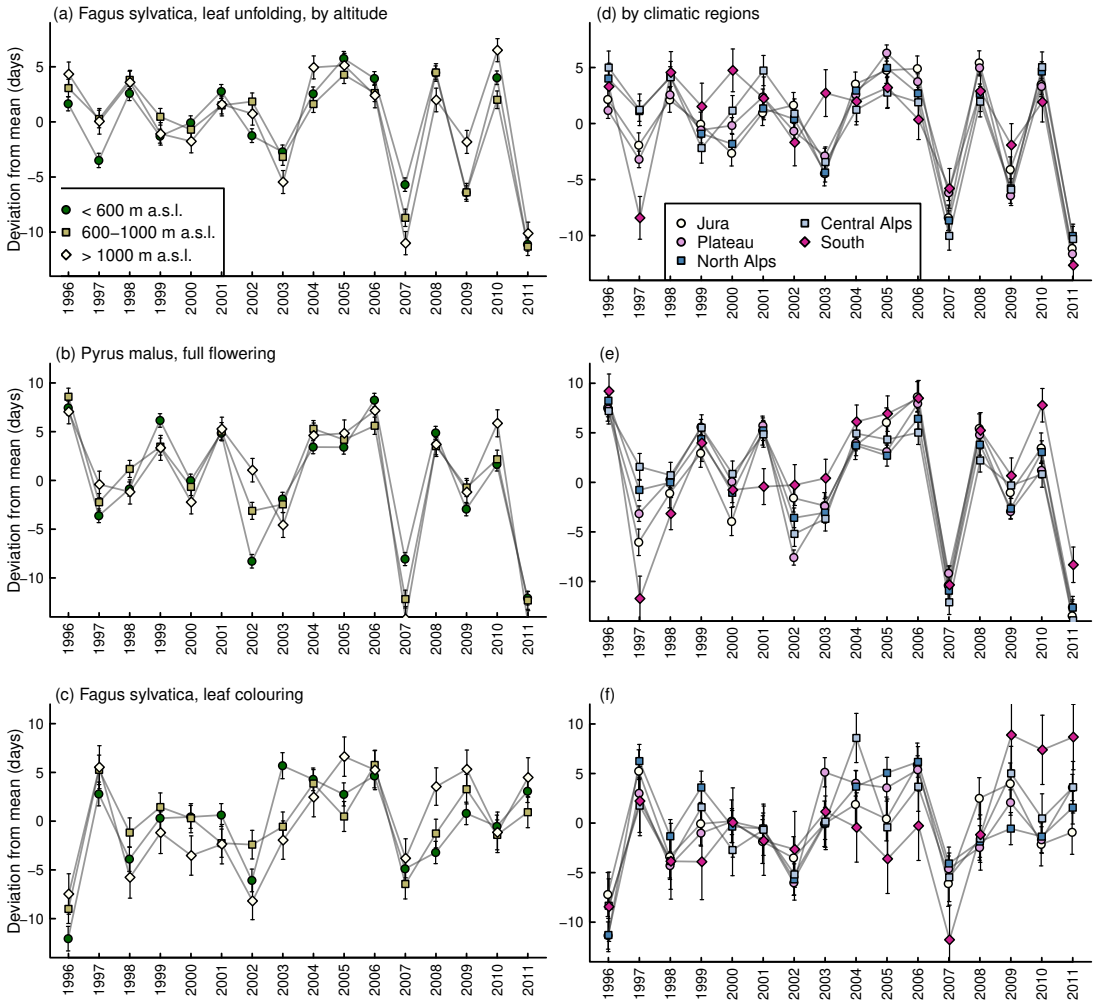


Figure 11: Interannual variation in the onset dates of three phenophases (a–c) in three altitudinal layers and (d–f) in five main climatic regions. Data represent the deviations from each station's mean date in 1996–2012. Mean deviations ($\pm$ se) per layer or region and year were determined with mixed models.

stations are virtually uncorrelated. For the other stages, the effective range ($\geq$ 192 km) includes most of the other stations as 86% of the pairwise distances are $<$ 192 km. However, a considerable nugget effect for all phases except for leaf colouring implies that even immediately adjacent stations are only weakly correlated. Semivariograms (not shown) also indicate that spatial correlation mainly reflects large-scale spatial trends, while most of the interannual variation occurs at a strictly local scale.

The inclusion of temporal autocorrelation substantially improves the models for all focal phenological stages, i.e. AIC values are considerably reduced compared to models without autocorrelation (Table 8). This reflects strong temporal autocorrelation, with coefficients ϕ of 0.4–0.7. Thus, while observations carried out at neighbouring stations show little dependence, observations carried out at the same station show strong dependence, which again highlights the importance of local-scale variation. With temporal autocorrelation, regression coefficients for annual means are estimated slightly more precisely (Table 8).

In conclusion, the magnitude of interannual variation is not consistently related to altitude, climatic regions or spatial distance, except for a slight trend towards higher variability of spring phenophases at

Table 7: Effect of including spatial correlation in models for annual deviations of phenological onset dates: AIC of models with and without spatial correlation, AIC difference, standard error of coefficients for annual deviations estimated from models with and without spatial correlation structure, effective range of the correlation function (3δ in km), and correlation of adjacent stations (distance $\downarrow$ 0). Models are fitted to the four focal phenological stages and include species-specific annual deviations. Spatial correlation of residuals is modelled by an exponential correlation function with nugget effect for observations grouped by species and year.

Stage	AIC			se of coefficients		range	correlation
	with	without	diff	with	without	(km)	$1 - c_0$
Leaf unfolding	118965	119271	–305	1.43	0.72	192.0	0.13
Full flowering	127194	127508	–314	3.04	1.04	365.1	0.20
Flowering of herbs	101176	101390	–214	3.05	1.28	240.6	0.16
Leaf colouring	93883	93944	–61	1.18	1.11	16.8	0.38

Table 8: Effect of including temporal autocorrelation in models for annual deviations of phenological onset dates: AIC of models with and without temporal autocorrelation structure, AIC difference, standard error of coefficients for annual deviations estimated from models with and without temporal autocorrelation structure, and autoregressive coefficient ϕ. Models are fitted to the four focal phenological stages and include species-specific annual deviations. Autocorrelation is modelled as AR1 process for observations grouped by station and year.

Stage	AIC			se of coefficients		ϕ
	with	without	diff	with	without	
Leaf unfolding	116309	119271	–2962	0.69	0.72	0.442
Full flowering	122713	127508	–4795	0.98	1.04	0.572
Flowering of herbs	95022	101390	–6368	1.14	1.28	0.698
Leaf colouring	91486	93944	–2458	1.05	1.11	0.473

higher altitude. Deviations of individual phenophases in individual years can present substantial altitudinal and regional differences as well as spatial correlation, but the patterns change from year to year. Spatial correlation mainly reflects large-scale spatial trends rather than similarities of neighbouring stations.

Long-term trends

Phenology has shifted towards earlier onset dates in 24 of the 26 phenophases recorded since 1970 (Fig. 12, Appendix 6). Linear trends for these phenophases are significant, i.e. 95% confidence intervals for the slope cover only negative values (Fig. 12, Appendix 6). Only leaf colouring and leaf drop in *Fagus sylvatica* show no trend (Fig. 12c). The trend towards earlier phenology is most pronounced for the flowering of woody species and fruit maturity, with an average slope of nearly 4 days/decade, followed by leaf unfolding and the flowering of herbs, with an average slope of about 2.5 days per decade (Table 9).

Trends vary considerably among stations, with standard deviations of 3–4 days per decade (Table 9). For almost all phenophases, some stations even show a trend towards later phenology, i.e. opposite to the general trend (see Appendix 12a). Long-term trends correlate negatively with altitude, meaning

that shifts towards earlier onset dates tend to be more pronounced at higher elevation (Table 9). The relationship between trends and altitude is plotted in Appendix 7 for each phenophase. The relationship is significant ($p < 0.05$) for 11 phenophases. Long-term trends show only weak spatial correlation (Table 9).

Altitude and climatic regions jointly explain on average 16% of the variation among stations in long-term trends of individual phenophases (range: 3–33%). On average, the effects of altitude (mainly local variation), climatic regions and their interaction explain a similar (small) fraction of variation (Table 10). However, the relative size of these effects differs considerably among individual phenophases (see the large standard deviations in Table 10). Fig. 13 shows two examples of phenophases whose long-term trends are significantly related to altitude and climatic regions, respectively.

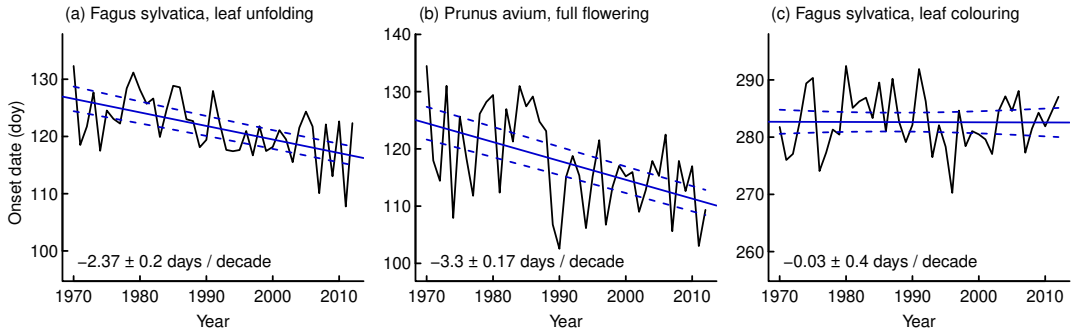


Figure 12: Long-term changes and linear trends (1970–2012) in onset dates of three selected phenophases. Regression lines (with 95% confidence intervals) are derived from mixed models including random effects of stations. Regression slopes ± se are given. Negative slopes indicate a trend towards earlier onset dates. The trend can be regarded as statistically significant ($p < 0.05$) if mean$+1.96 \cdot$se< 0. Similar graphs for all phenophases are found in Appendix 6.

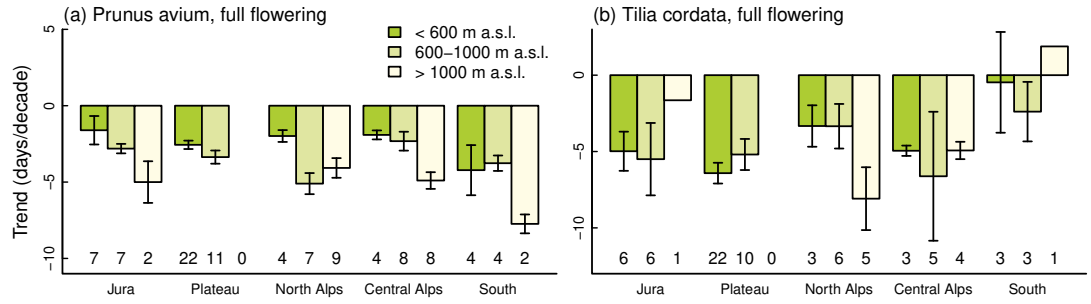


Figure 13: Differences in phenological trends 1970–2012 among altitudinal layers and climatic regions for two selected phenophases with (a) trend depending on altitude and (b) trend differing among regions. Trends (regression slopes against years) were calculated for each station; bars show means ± se per altitudinal layer and climatic region; the number of stations is given below each bar.

Table 9: Phenological trends 1970–2012 in relation to altitude and spatial distance: mean trends (slopes from linear regression of onset dates against years) and standard deviation of trends between stations, correlation of trends with the altitude of stations (Pearson's *r* and spatial correlation (Mantel *r*). Statistics were calculated for each phenophase. Means (and sd) of each phenological stage are given in the table, *n* is the number of species per stage.

Stage	*n*	mean trend (days/decade)	sd of trends (days/decade)	correl. with altitude (*r*)	spatial correl. (Mantel *r*)
Leaf unfolding	5	–2.39 (0.50)	2.91 (0.64)	–0.257 (0.080)	0.028 (0.049)
Full flowering	9	–3.93 (0.69)	3.16 (1.28)	–0.171 (0.183)	0.047 (0.036)
Flowering of herbs	6	–2.75 (1.02)	3.28 (0.81)	–0.144 (0.119)	0.050 (0.017)
Fruit maturity	2	–4.28 (1.70)	4.22 (2.11)	–0.034 (0.103)	0.087 (0.026)
Leaf colouring	2	–1.31 (1.89)	4.22 (0.45)	–0.092 (0.197)	0.001 (0.041)
Leaf drop	2	–0.57 (0.74)	3.43 (0.75)	–0.212 (0.011)	0.021 (0.052)

3.3 Discussion

Number and distribution of phenological stations

The overall result apparent from this analysis is that altitude is the main factor structuring phenological variation across Switzerland. Strong relationships between mean onset dates of spring phenophases and altitude were expected as they have been described many times before (e.g. *Dittmar and Elling*, 2006; *Pellerin et al.*, 2012). Not only mean onset dates, but also the variability of onset dates and the slope of long-term trends appear to depend on altitude for some of the phenophases. Accordingly, a good representation of the altitudinal range of the observed plant species is critical for the representativeness of the Swiss Phenology Network.

Besides altitude, large-scale spatial patterns also exist, which often (but not always) correspond to the climatic regions. For example, leaf unfolding and flowering of some woody species differ by more than 10 days between the earliest region (central Alps) and the latest region (Jura). In addition, changes with altitude can differ among regions. Hence, both altitude and climatic regions should ideally be represented sufficiently in the SwissPhenology Network.

If we consider the current numbers of stations per region and altitudinal layer (Table 11), it appears that high elevations in the Jura and southern Switzerland and low elevations in the central Alps are less represented. Although this partly reflects the size and topography of each region, it can serve as an indication which locations deserve priority in maintaining and possibly expanding the Swiss Phenology Network. Since spatial patterns in phenology may differ from the climatic regions (e.g. Fig. 8a, b), a sufficient density of stations in every part of the country should be targetted. The quasi-absence of spatial dependence between neighbouring stations implies that efforts to include new stations in a region are not constrained by the need to ensure a certain distance to existing stations. Although

Table 10: Effects of altitude and regional differences on phenological trends 1970–2012 (slopes from linear regression of onset dates against years), described by the percentage of variation explained by each factor and their interaction. Altitude varies both within regions ('local') and among regions ('regional'). Means (and sd) of the phenophases belonging to each stage are given; *n* is the number of phenophases (i.e. species) per stage. Superscript numbers indicate the number of phenophases for which the effect is significant ($p < 0.05$). See Table 3 for further details.

Stage	*n*	altitude		region	interaction	residual
		local	regional			
Leaf unfolding	5	3.3 (2.3)^1	0.7 (1.6)	3.4 (1.6)^0	7.5 (7.1)^1	85.2 (6.4)
Full flowering	9	6.0 (7.4)^3	–0.0 (3.4)	6.0 (4.1)^2	5.8 (2.8)^1	82.2 (9.1)
Flowering of herbs	6	3.0 (4.0)^2	–0.7 (2.3)	6.4 (3.1)^1	7.4 (6.3)^2	84.0 (8.3)
Fruit maturity	2	5.4 (7.2)^1	–2.9 (3.9)	8.5 (0.4)^0	4.3 (1.0)^0	84.7 (2.7)
Leaf colouring	2	10.5 (11.5)^1	–2.5 (5.7)	6.7 (3.5)^0	0.5 (0.6)^0	84.9 (8.7)
Leaf drop	2	4.4 (5.9)^1	3.6 (1.4)	5.9 (2.3)^0	2.8 (1.0)^0	83.3 (8.6)

spatial dependence almost certainly exists at a local scale, the range of this dependence is probably too small to be a constraint in practice.

How many stations are needed? This question can only be answered with reference to specific statistical targets such as an acceptable error in the estimation of a certain parameter or the required statistical power in testing a difference. A few arbitrary examples shall be considered here.

We might wish to estimate mean phenological onset dates per altitudinal layer with a certain error tolerance. Assuming normality, the width of a 95% confidence interval for the mean is $w = 2 \cdot t_{0.975,n-1} \cdot s/\sqrt{n}$, where $t_{0.975,n-1}$ is the quantile of the t distribution with $n-1$ degrees of freedom, s the sample standard deviation, and n the number of stations. In Table 12 these numbers are given for 95% confidence intervals of 10 days or 5 days wide. With the currently available number of stations, mean onset dates per layer can be estimated precisely (with narrow confidence intervals) for leaf unfolding and the flowering of woody species but less precisely for the flowering of herbs and fruit maturity. Results in Table 12 refer to mean onset dates over 17 years. For individual years, standard deviations are on average 1.26 times larger, hence 1.6 times more stations would be required.

To study phenological variability, such as responses to extremely warm springs, deviations of particular years from long-term means are of interest. Table 13 gives the residual standard deviation of models for interannual variation of the four focal phenological stages (mixed model (1) fitted separately to each altitudinal layer) and average standard errors of estimated coefficients for annual deviations. Based on the size of standard errors, estimates for altitudes above 1000 m a.s.l. are up to 3 times less precise than for the lowlands due to both a smaller number of stations and higher residual variation (Table 13). A similar analysis for climatic regions (not shown) indicates larger estimation error for the Jura and southern Switzerland, which could be a further incentive to increase their representation in the network.

Table 11: Number of active stations in the SwissPhenology Network per region and altitudinal layer: Stations included in this analysis, i.e. with $\geq$ 14 years of data, and total number of stations.

Region	stations included in this analysis			all active stations		
	$<$ 600 m	600–1000 m	$>$ 1000 m	$<$ 600 m	600–1000 m	$>$ 1000 m
Jura	6	8	5	10	8	5
Plateau	38	10	0	51	13	0
North Alps	9	11	13	10	12	15
Central Alps	4	6	15	4	9	16
South	5	5	3	7	5	4

Table 12: Standard deviation (sd) of mean phenological onset dates between stations within each of three altitudinal layers and number of stations needed to obtain 95% confidence intervals (CI) of a certain width for the mean date of a variable in a layer. All calculations are based on mean dates 1996–2012 of each variable at each station. Standard deviations were calculated for each variable but means per phenological phase are given in the table, and these means are used to derive the required number of stations. A confidence interval of 10 days corresponds to mean $\pm$ 5 days. Altitudinal layers and the number of stations included in this analysis are L: $<$ 600 m (n = 62), I: 600–1000 m (n = 40), H: $>$ 1000 m (n = 36).

Stage	sd per layer (days)			number of stations needed for... CI = 10 days			CI = 5 days		
	L	I	H	L	I	H	L	I	H
Leaf unfolding	6.3	7.0	6.9	7	9	8	28	35	34
Flowering start	8.2	10.5	10.4	12	19	19	48	78	77
Full flowering	8.0	9.8	10.6	11	17	19	46	68	80
Flowerig herbs	10.6	11.0	13.6	19	21	32	80	86	131
Fruit maturity	16.2	12.6	16.3	45	27	45	185	112	188
Leaf colouring	9.3	10.2	9.6	15	18	16	61	74	65
Leaf drop	8.3	9.1	8.8	12	14	14	49	59	55

Table 13: Estimation errors for annual deviations from the mean: residual standard error of models for each phenological stage and altitudinal layer (error for individual observations), and average standard error of coefficients for annual deviations (error for annual means per altitudinal layer). LU = leaf unfolding, FF = Full flowering, LC = Leaf colouring, FH = flowering of herbs.

Altitudinal Layer	n	residual standard error				standard error of coefficients			
		LU	FF	LC	FH	LU	FF	LC	FH
$<$ 600 m	62	6.9	9.2	11.2	11.9	1.6	1.8	2.7	2.3
600–1000 m	40	7.2	10.2	10.9	11.6	1.8	2.8	2.8	3.3
$>$ 1000 m	36	8.0	11.7	10.1	13.9	2.3	5.4	3.0	5.1

Temporal trends

Phenological trends during the time period 1970–2012 were analysed with linear regression, as has been done in many previous studies (e.g. *Menzel et al.*, 2001, 2006; *Estrella et al.*, 2009; *Defila and Clot*, 2001). Earlier analyses were mostly based on data until 2000. The extension of the time period until 2012 shows that spring phenology has continued to shift forward for many variables. This continued advance is largely due to the extremely early years 2007 and 2011 (*Rutishauser et al.*, 2008; *Maignan et al.*, 2008). The longer the time period considered, the more questionable is the use of simple linear models to describe the trends (*Dose and Menzel*, 2004). Obviously, linear trends cannot continue indefinitely; they must be restricted to certain time periods and eventually level off or reverse. As a result, statements about the slope or significance of linear trends depends critically on the time period considered (*Roetzer et al.*, 2000; *Dose and Menzel*, 2004). The period 1970–2012 has the advantage that the years with most rapid warming (1985–2000) are just in the middle, so that associated trends can be estimated more reliably than if the change occurs at the very end of the observation period (*Dose and Menzel*, 2004).

At least for the period considered here (1970–2012), the slope of linear trends can be estimated quite precisely with the current set of stations in the Swiss Phenology Network. Standard errors of the slopes for spring phenophases range between 0.15 and 0.45 days/decade, i.e. errors are much smaller than the estimated values.

Results also suggest that temporal trends of some phenophases (particularly leaf unfolding and flowering of trees) depend on altitude, being stronger at high elevations. This result contrasts with earlier ones at European scale, where there was a tendency for weaker trends at high elevation (*Menzel et al.*, 2001, 2006; *Schleip et al.*, 2009). However, relationships with altitude were generally very weak (*Estrella et al.*, 2009) and may have been driven by associations between altitude, latitude and continentality. For example, *Menzel et al.* (2006) found weaker negative trends at higher altitudes across Germany for several phenophases, but higher altitudes are concentrated in the southern part of the Germany; negative trends tended to be stronger in the northern parts of the country, so that the relationship with altitude might just be coincidential. In studies focusing on Alpine regions, *Defila and Clot* (2001) and *Ziello et al.* (2009) found more negative trends at higher altitude but noted that the relationship is weak. *Vitasse and Basler* (2013) found more negative trends at higher altitude for beech (*Fagus sylvatica*) and suggested that this result indicates a photoperiodic requirement for spring development in this particular species. The present work finds stronger trends at high elevation for many species and for different phenological stages. A possible explanation for the stronger trends at high elevation is that temperatures in April and May (when spring starts at high altitudes) have increased much more over the past 40 years (0.7–0.8 °C/decade) than temperatures in January and February (0–0.3 °C/decade), while the temperature trend in March was intermediate (0.55 °C/decade). Thus, plants growing at higher altitude and developing later in spring experienced a stronger warming trend during their spring season than plants growing in the lowlands.

Not only long-term trends but also interannual variation tends to increase with altitude in some of the phenophases. This also seems to have a simple climatic explanation: Interannual variation of February and March temperatures increases with altitude. For example, for February, the standard deviation of

monthly mean temperatures between 1996 and 2012 is 3.5–4.5 °C for most stations below 700 m a.s.l. but 4.5–5.5 °C for most stations above 700 m a.s.l.

Spatial trends

Spatial variation is dominated by large-scale trends, which are not necessarily reflected by climatic regions. Maps of spatial fields are a natural way of representing these trends (*Schleip et al.*, 2009). The maps derived from mean onset dates (Fig. 8) reveal clear differences among phenophases. Because of the dominant influence of altitude on phenology, such maps are most informative if they separate the fields describing effects of altitude and deviations from the altitudinal trend. The quasi-absence of small-scale spatial dependence means that predictions of phenological onset dates for a new location can be derived from these large-scale fields (though with considerable uncertainty!) while there is little point in interpolating dates between neighbouring stations.

Temporal trends in phenology can be modelled with spatial models in the same way as mean onset dates. *Schleip et al.* (2009) compared the spatial fields of eight spring phenophases across Europe and found that spatial fields of mean onset dates are all similar while spatial fields of temporal trends differ considerably among phenophases. Modelling interannual variation is more challenging because spatial patterns of stations with large and small deviations from the mean appear to vary from year to year. This space-time interaction would have to be modelled in an appropriate way, e.g. in a multivariate approach (*Studer et al.*, 2005).

4 Similarities of phenological stations

4.1 Aims and methods

The previous chapter has related phenological variation to factors associated with the spatial location of stations. We now consider similarities between stations irrespective of their location. As before, such an analysis can be based on means of each time series, interannual variation or long-term trends. This chapter focuses on interannual variation, which provides the most detailed information. Results for the two other levels are briefly mentioned in the discussion part.

Correlations. Stations are compared based on the time series from 1996–2012. Correlations of these time series are calculated for all pairs of stations and for each phenophase provided that data are available for ≥ 6 years. Spearman rank correlations are used to reduce the variability of results obtained from short time series. This yields a correlation matrix of stations for each phenophase. Correlation matrices for individual phenophases are then combined to matrices of mean correlation coefficients for the four focal stages. The four combined correlation matrices are used to determine how many stations are, on average, well correlated to any particular station, considering only positive correlations: For one phenological stage, this is the mean number of coefficients ≥ 0.6 or ≥ 0.7 per column of the corresponding correlation matrix. For two or more stages simultaneously, this is the mean number of coefficients that are ≥ 0.6 or ≥ 0.7 simultaneously in two or more correlation matrices.

Clustering. To identify groups of stations with similar patterns of interannual variation, hierarchical agglomerative clustering (complete linkage) is performed both with correlation matrices for individual phenophases and with the combined correlations for the four focal stages. Only positive correlations indicate station similarity and should be the basis for clustering. Therefore, negative correlations are set to 0 before taking the complement ($d = 1 - r$) to obtain a distance matrix for cluster analysis. Stations with any missing correlations must be excluded before cluster analysis. This is done sequentially, one station at the time, removing the station with the largest number of missing correlations at each step. Thus, stations with only few missing correlations can remain in the analysis if the partner(s) of missing correlations are excluded. Only 4–21 stations are excluded from the combined correlation matrices, while up to 87 stations are excluded from the correlation matrices for individual phenophases. The results of cluster analysis are evaluated visually by checking the grouping structure apparent in dendrograms. To see whether groups of stations identified by cluster analysis correspond to one of the spatial classifications, information about altitudinal layers and climatic regions is added to the dendrograms.

Representative stations. To identify stations that are most representative of the overall patterns of interannual variation, time series of individual stations are correlated with the average time series for

each phenophase. Average time series are derived from the coefficients of mixed models with fixed effect of year and random effect of station. A matrix of correlations is obtained, where rows represent stations and columns represent phenophases. Correlations are combined by taking row means for each phenological stage, leading to a matrix with four columns for the four focal stages. The highest and lowest correlations in this matrix (overall or per column) are used to identify representative stations (most correlated with the average time series) and deviant stations (least correlated with the average time series). The same analysis is also performed separately for each of the five main climatic regions, i.e. all stations from one region are correlated with the mean time series of that region to identify stations that are most or least representative of their region.

4.2 Results

Correlations between stations

Correlations between the time series of different stations range from almost –1 to +1 for each of the phenophases, with mean correlations (for all pairs of stations) of 0.14–0.60 for leaf unfolding or flowering and 0.09–0.19 for leaf colouring. Thus, on average, time series correlate only weakly to moderately between stations, although some pairs of stations are strongly correlated for each of the phenophases. These strongly correlated pairs of stations differ from phenophase to phenophase. Even within the same phenological stage, correlations between stations obtained for different species are only weakly related to each other.

For each pair of stations, mean correlations per phenological stage are obtained by taking the mean of correlations obtained for individual species. On average (over all pairs of stations), these mean correlations are 0.36 for leaf unfolding, 0.44 for full flowering, 0.14 for leaf colouring, and 0.35 for the flowering of herbs (Table 14). Mean correlations for one stage are again only weakly related to mean correlations for another stage, i.e. pairs of stations with strongly correlated time series for one stage do no necessarily have strongly correlated time series for another stage (Table 14). Mean correlations above 0.6 are relatively frequent for individual stages, i.e. most stations correlate at least so strongly with at least one other station for a single stage (upper part of Table 14). Correlations above 0.7 are much rarer (upper part of Table 14). Furthermore, because correlations for different stages are weakly related, only few pairs of stations correlate well with each other for two or more stages simultaneously (lower part of Table 14).

Clustering of stations

Hierarchical clustering of stations based on correlation matrices for individual phenophases (Appendix 8) generally shows pairs or small groups of strongly correlated stations. Larger groups of stations are apparent only for some of the phenophases, only for some of the stations, and usually only with moderate correlations. As noted above, the identity of strongly correlated stations differs among phenophases.

Clustering of stations based on mean correlations per stage (Fig. 14) also shows a weak grouping structure. Any major groups are clustered at low correlation levels (e.g. Fig. 14b). Even if we form as

Table 14: Correlations between the time series of stations (1996–2012) obtained for different phenological stages. Correlations calculated for each phenophase were averaged to obtain mean correlations per stage. The first column in the table gives mean correlations per stage over all pairs of stations (upper part of table) and how they correlate between stages, i.e. to what extent pairs of stations with strongly correlated time series for one stage also have strongly correlated time series for another stage (lower part of table). Based on mean correlations per stage, pairs of stations with 'well correlated' time series were identified, using $r \geq 0.6$ or $r \geq 0.7$ as threshold. The second and third column in the table give the mean number of stations well correlated to a particular station considering one, two or all stages (the threshold must be fulfilled for all stages considered). Numbers smaller than 1 indicate that many stations are not well correlated to any other station.

Stages	mean r and correlation of r	mean number of similar stations at $r \geq 0.6$	at $r \geq 0.7$
Leaf unfolding	0.36	8.51	1.57
Full flowering	0.44	24.70	7.41
Flowering of herbs	0.35	9.52	1.96
Leaf colouring	0.14	1.88	0.68
Leaf unfolding + Full flowering	0.34	4.16	0.41
Leaf unfolding + Flowering of herbs	0.30	1.68	0.07
Leaf unfolding + Leaf colouring	0.07	0.20	0.03
Full flowering + Flowering of herbs	0.34	3.75	0.33
Full flowering + Leaf colouring	0.04	0.36	0.03
Leaf colouring + Flowering of herbs	0.07	0.17	0.01
All stages		0.00	0.00

many as 30 clusters, some stations belonging to the same cluster correlate as weakly as $r = 0.35$ for leaf unfolding, $r = 0.43$ for full flowering, $r = 0.29$ for flowering of herbs, and $r = 0.12$ for leaf colouring. If we combine these 30-cluster classifications for two stages, such that stations are grouped only if they cluster together for both stages, 59–85% of the stations form single-station clusters, and most of the remaining stations form pairs; there are only very few clusters of three or more stations. If we combine classifications for the four stages, all stations form single-station clusters.

Clusters occasionally include stations with similar altitude and from the same or neighbouring climatic regions (see coloured points in Fig. 14), but more often this is not the case, and clusters consist of stations from diverse locations. Stations belonging to relatively large clusters (≥ 6 stations) do not form spatial groups when plotted on a map (not shown).

Representative stations

If the time series of individual stations are compared to the mean time series of a phenophase, some stations strongly correlate with the mean time series, and others strongly deviate from it (Fig. 15). The proportion of strong and weak correlations differs considerably among phases: Many stations strongly correlate with mean time series for leaf unfolding and full flowering (Fig. 15a, b) while many stations correlate weakly or even negatively with mean time series for leaf colouring (Fig. 15c). Correlations are distributed rather homogeneously in space over Switzerland, although the most representative stations

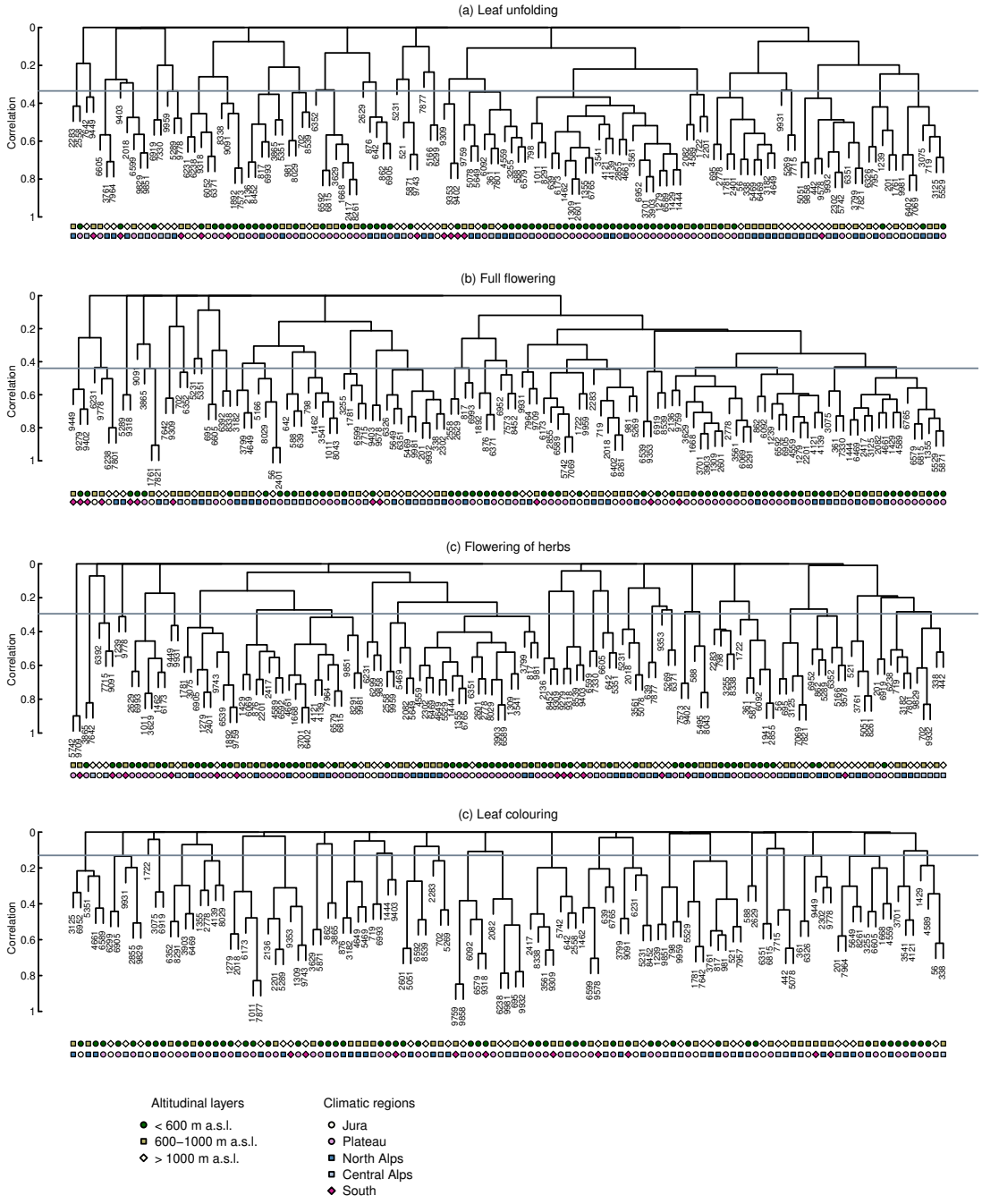


Figure 14: Dendrograms from cluster analysis of stations based on the mean correlations of time series for the four focal phenological stages. Horizontal lines indicate the correlation at which the dendrogram must be cut to obtain 30 groups of stations. Symbols below dendrograms indicate the altitude and region of stations. Series of identical symbols would indicate that clusters reflect the corresponding classification.

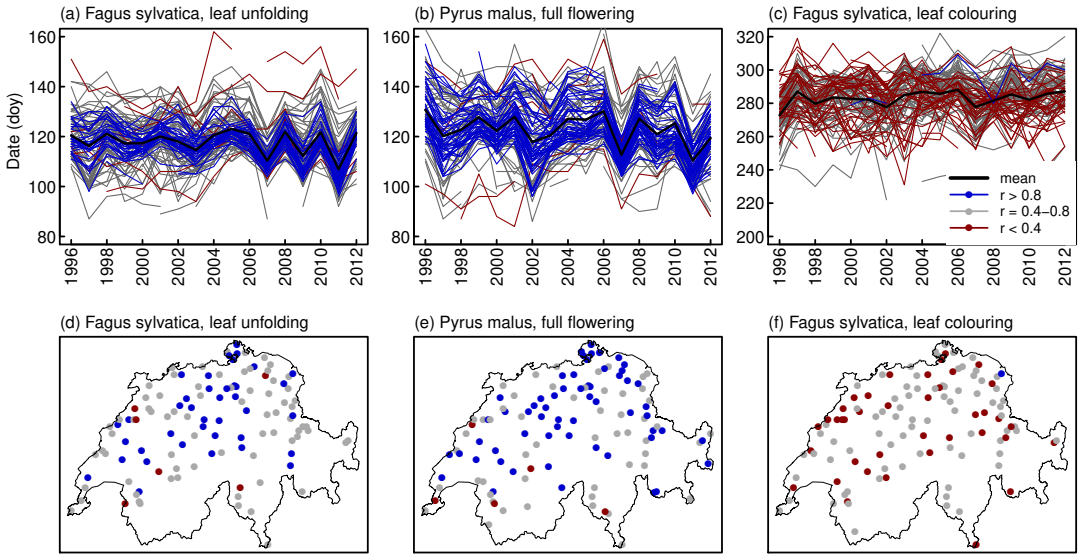


Figure 15: Illustration of correlations between the time series of individual stations and the overall mean time series (as a measure of the stations' representativeness) for three phenophases. In (a)–(c), lines represent each station's time series, and in (d)–(e) points show their spatial distribution. Colours indicate the degree of correlation with the mean time series according to the legend in (c).

tend to be concentrated in the Plateau (Fig. 15d–f).

The distribution of correlations for all phenophases or for those belonging to one phenological stage (Appendix 9A) reveals stations that are well correlated with the mean time series (representative stations) and stations that are weakly correlated with it (deviant stations, Table 15). The identity of the representative stations tends to be different for each phenological stage, while deviant stations often appear for several stages (Table 15). However, most stations show a broad range of correlations even for one phenological stage, being well correlated with the mean time series of some species and poorly correlated with others (Appendix 9A). These conclusions also hold if calculations are done separately for each of the climatic regions (Appendix 9B, Table 15).

4.3 Discussion

The main result is that time series from 1996–2012 mostly correlate weakly between stations, while grouping structures revealed by cluster analysis are both weak and inconsistent. In addition, even neighbouring stations correlate hardly more than any other pair of stations does on average, consistent with the weak small-scale dependence described in Chapter 3. These results contrast with those obtained for the pollen monitoring network, which exhibits two rather clear and consistent clusters in the eastern and western Plateau, respectively (*Gehrig*, 2012). The main reason for this difference certainly is that pollen traps capture pollen from many plants growing at some distance around the trap, whereas phenological observations concern a single plant individual or population, so that genetic variation and variation in local site conditions have a strong influence (e.g. *Wielgolaski*, 2001). In addition, pure observation error is likely to be relatively greater for phenological observations than for pollen counts (*Beaubien and Hamann*, 2011).

Table 15: Stations that are most representative and stations that deviate most from the overall patterns of interannual variation in the entire country and in each of the climatic regions. Representativeness and deviations are measured by the correlation between the time series (1996–2012) of a station and the mean time series of the country or region, calculated for each phenophase and then averaged over species for each phenological stage. The four stations with the highest and lowest mean correlation are given for each phase and region (ordered by decreasing or increasing correlation, respectively). See Appendix 1 for information about stations.

Region	leaf unfolding	full flowering	flowering of herbs	leaf colouring
Most representative stations				
Country	1309, 1444, 4649, 3701	521, 5742, 3903, 1309	1892, 2855, 1355, 5529	5289, 9743, 56, 817
Jura	6069, 1668, 6238, 6231	6402, 6052, 1941, 6069	6952, 6351, 1781, 1668	1781, 1892, 6402, 6352
Plateau	1309, 1444, 2601, 1279	3903, 1309, 3701, 1444	3701, 2855, 5529, 1355	6579, 6173, 6599, 8338
Norther Alps	4649, 2302, 3125, 5469	6469, 4649, 3125, 2302	7964, 4649, 7821, 3182	5289, 817, 7964, 3761
Central Alps	338, 201, 9851, 442	521, 9829, 338, 9932	702, 9932, 338, 7069	56, 695, 9829, 9932
Southern Switzerland	9279, 9402, 9353, 9578	9578, 9709, 9279, 9403	9279, 9318, 9743, 9578	9318, 9743, 9759, 9353
Most deviant stations				
Country	521, 7877, 8043, 5231	6299, 7642, 5231, 3865	7957, 9091, 9353, 5231	6919, 7573, 9091, 9931
Jura	6352, 6392, 8539, 1892	6299, 6392, 6952, 1722	6052, 8539, 1722, 6392	1722, 6299, 6069, 6351
Plateau	8043, 2629, 3865, 8338	3865, 8338, 2629, 2558	8043, 3865, 6919, 6605	6919, 6815, 3865, 8291
Northern Alps	7877, 5289, 5231, 862	5231, 5351, 5289, 7877	7957, 5231, 2283, 7877	5051, 2302, 5231, 5495
Central Alps	7642, 521, 7573, 642	7642, 702, 9851, 639	7573, 9931, 7642, 798	7573, 9931, 7642, 702
Southern Switzerland	9449, 9403, 9091, 9709	9309, 9091, 9449, 9743	9709, 9091, 9778, 9449	9091, 9402, 9403, 9449

In the context of climate change research, there is a growing interest in phenological responses to climatic extremes, such as cold and hot spells (*Menzel et al.*, 2011; *Maignan et al.*, 2008), extended drought or extreme rainfall (*Jentsch et al.*, 2009). The present results suggest that the responses to those events are likely to vary from station to station, and that this variability cannot be reduced by grouping stations according to altitude or climatic regions or the results of cluster analysis. Instead, the large number of stations included in the Swiss Phenology Network leads to precise estimates of annual deviations (e.g. Fig. 9) despite the underlying local variation.

In this chapter, similarities between stations have been measured by the correlations of their time series, i.e. they are based on patterns of interannual variation. Of course, similarities between stations can also be analysed based on mean onset dates or based on long-term trends. In this case, time series are summarized by a single value per station and phenophase, so that stations can be directly compared for all phenophases together. The main results (not shown in this report) can be summarized in a few words: Based on mean onset dates, cluster analysis reveals three groups that mostly correspond to the three altitudinal layers ($<$ 600 m, 600–1000 m, $>$ 1000 m). Based on long-term trends, cluster analysis reveals two groups with different trends for leaf colouring and leaf drop in *Aesculus hippocastanum* and *Fagus sylvatica*. While onset dates of these autumn phenophases advanced over the last four decades at most stations, a smaller group of scattered stations showed the opposite trend.

5 Phenological responses to temperature

5.1 Aims and methods

This chapter analyses phenological responses to recent temperature fluctuations. It examines how variable these responses are, whether they depend on the altitude and region of stations, and whether estimates of temperature sensitivity obtained from recent fluctuations correspond to those estimated from older data, long-term trends or altitudinal gradients. Precise and consistent estimates of temperature sensitivity are essential for predicting phenological responses to future climate warming, and any factors influencing temperature sensitivity may have to be included in predictive models.

Temperature data from the SwissMetNet network (automatic surface observation network) are used in this analysis because temperatures are not recorded directly at the phenological stations. Each phenological station is manually matched with one of the SwissMetNet stations providing homogenized temperature time series back to 1996 for the recent data and back to 1970 for the long-term data. Matching is primarily based on geographic coordinates, and secondarily on altitude and topographic position (e.g. mountain or valley). Temperature data are daily mean temperatures at 2 m above soil, adjusted for the difference in altitude between phenological and climatic stations. The adjustment is based on altitudinal gradients in temperature derived from the 1981–2010 norm values of monthly mean temperatures at 89 stations[3]. Because temperature decreases non-linearly with altitude, different slope coefficients are used for matched stations with mean altitude above and below 700 m a.s.l. (Appendix 10).

To separate temperature fluctuations from spatial variation in temperatures, the adjusted temperature time series associated with each phenological station are converted into deviations from a station's mean over the time period considered. Deviations are calculated separately by station and by day of the year, so that the mean deviation of each station is 0 for each day of the year. Once temperature time series have been associated with each phenological station, a temperature (or set of temperatures) must be assigned to each phenological observation. Previous studies often found that mean temperatures in the 1–3 months preceding a phenophase correlate well with the onset dates (*Chmielewski and Rötzer*, 2001; *Menzel et al.*, 2006; *Ibáñez et al.*, 2010; *Sparks et al.*, 2009; *Vitasse et al.*, 2013). Therefore, for each phenophase at each station, the mean onset date (doy) over the study period (1970–2012 or 1996–2012) is determined, and mean temperatures of the 60 Julian days preceding this date are associated with each year's phenological observation. These temperatures are hereafter called 'local temperatures'. Responses to temperature fluctuations are calculated separately for the period 1996–2012 (recent data) and 1970–2000 ('old' data).

[3] http://www.meteoschweiz.admin.ch/files/kd/normwerte/norm8110/nvrep_tre200m0_de.txt

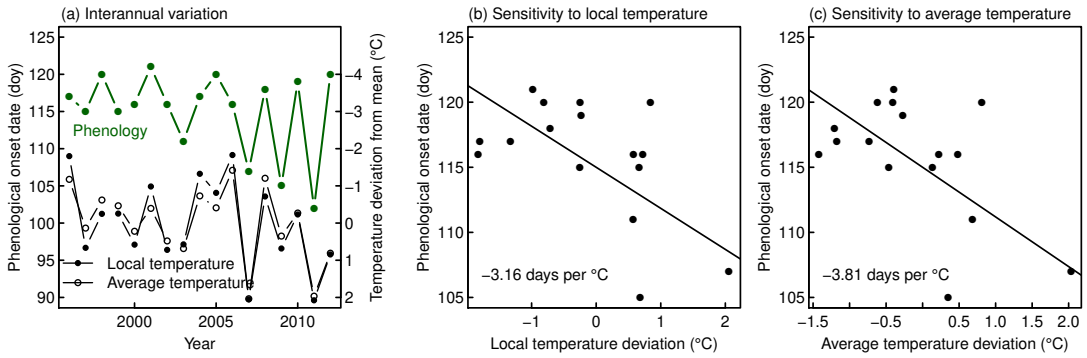


Figure 16: Determination of temperature sensitivity of a phenophase based on interannual variation between 1996 and 2012. (a) Time series of onset dates for leaf unfolding of *Fagus sylvatica* at station Zürich MeteoSchweiz together with local and average temperature deviations from the mean at this station. (b,c) Linear regression of onset dates on temperature deviations; the slopes are defined as temperature sensitivity.

For comparison, recent temperature fluctuations are also calculated in a simpler way that does not require station-specific temperatures: A single time series of average monthly mean temperatures is obtained by taking the mean of ten freely available homogenized time series of monthly mean temperatures[4]. Temperatures of each month are converted into deviations by subtracting the 1996–2012 mean for this month. For each phenophase and for each of the three groups of stations formed by cluster analysis (Chapter 2.2), a median onset date is determined. Temperature deviations of the two months preceding and including the median onset date are associated with each year's phenological observation; these are called 'average temperatures'.

Temperature sensitivity (the average difference in onset date per difference in temperature) is calculated for each phenophase and station as the slope of a linear regression of onset dates against local (recent and old) or average temperature deviations. The procedure is illustrated in Fig. 16.

For the trend 1970–2012, temperature sensitivity is defined as the ratio between the mean annual change in onset dates (regession slope of onset dates against years) and the mean annual change in temperature (regression slope of local temperature deviations against years). For altitudinal gradients, the slope of a linear regression of each station's mean onset date against altitude is converted into temperature sensitivity assuming a single thermal gradient of 0.6 °C per 100 m, which corresponds well to the overall temperature gradient in March-May (Appendix 10).

5.2 Results

Interannual variation in onset dates of spring phenophases (leaf unfolding and flowering) correlates negatively with temperatures in the two months preceding the average date of the phenophase (Table 16). The correlation with local temperatures is slightly stronger than the correlation with average temperatures. Fruit maturity correlates only weakly with temperature, and for autumn phenophases (leaf colouring and leaf drop), the correlation is weakly positive (Table 16). On average 31–40% of the interannual variation in spring phenology of woody plants and 4–24% of the variation in the other phases can be accounted for by temperature (Table 16).

[4] http://www.meteoschweiz.admin.ch/web/en/climate/climate_today/homogeneous_data.html

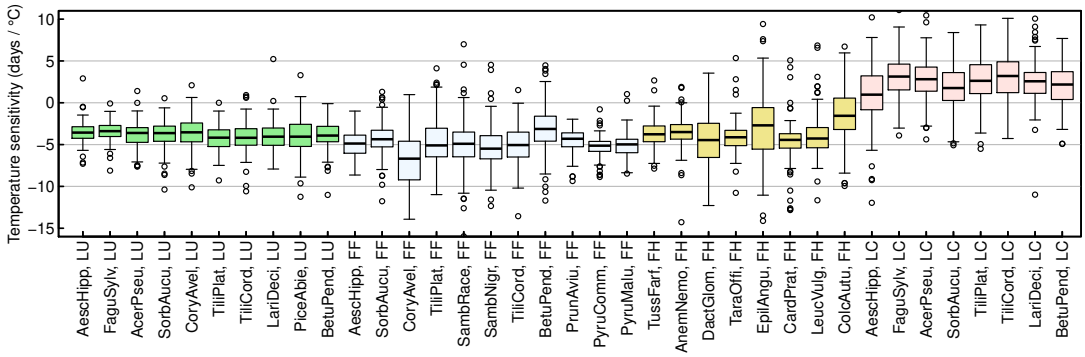


Figure 17: Distribution of temperature sensitivity coefficients based on year-to-year fluctuations in local temperatures between 1996 and 2012. Sensitivity coefficients were determined for each phenophase and station. Each box in the graph represents the distribution over stations for one phenophase.

Temperature sensitivity, i.e. the difference in phenological onset dates associated with a 1 °C difference in temperature, is highest for spring phenophases (Table 16). Temperature sensitivity coefficients vary considerably among stations, whereas average values are similar in different species for a given phenological stage, especially for leaf unfolding (Fig. 17).

On average, temperature sensitivity correlates only very weakly with the altitude of stations (Table 16). However, significant correlations ($p < 0.05$) exist for a few phenophases. Temperature sensitivity increases with altitude (more negative coefficients at higher altitude) for leaf unfolding of *Acer pseudoplatanus* and *Tilia cordata* as well as for the flowering (start and full) of *Aesculus hippocastanum*, *Prunus avium*, *Anemone nemorosa* and *Cardamine pratensis*. Temperature sensitivity decreases with altitude (more negative coefficients in the lowlands) for the flowering of *Corylus avellana*. Temperature sensitivity differs little among climatic regions, except for the flowering of *Corylus avellana*, whose greater temperature sensitivity in the Plateau just reflects the relationship with altitude (details not shown).

Different calculations of temperature sensitivity generally produce similar results (Fig. 18, Appendix 11). In particular, responses to recent fluctuations in average temperatures closely resemble those to fluctuations in local temperatures regarding both means and standard errors (Fig. 18a). The same holds for reponses to temperature fluctuations between 1970 and 2000; again, means and standard errors are similar to those for responses to recent temperature fluctuations (Fig. 18b).

Reponses to temperature trends between 1970 and 2012 are more variable than responses to recent fluctuations (Fig. 18c). This is due to the combined variability of linear trends in phenology and in temperature (Appendix 12). Temperature sensitivity coefficients are most variable for autumn and early spring phases because temperatures during this period showed only a weak warming trend (Appendix 12). These 'unreliable' sensitivity coefficients are generally more negative than the coefficients derived directly from temperature fluctuations, whereas the other coefficients correspond relatively well to each other (Fig. 18c).

Responses to altitudinal gradients in temperature are generally similar to those derived from temperature fluctuations for leaf unfolding and flowering, but more negative for fruit maturity and less positive for leaf colouring (Fig. 18c).

Table 16: Temperature sensitivity of phenology derived from interannual variation between 1996 and 2012. Mean correlation (Pearson's *r*) between phenological onset dates and local or average temperatures (T, mean temperature of the 2 months before the phenophase), fraction of variation explained by local temperatures (adjusted r^2 from linear regression), mean temperature sensitivity at all stations, and correlation of temperature sensitivity with the altitude of stations. All statistics were calculated for each phenophase; means (and sd) per phenological stage are given in the table.

Stage	association of phenology with temperature			temperature sensitivity	
	correl. with local T	correl. with average T	adjusted r^2 local T	mean (days/°C)	correl. with altitude
Leaf unfolding	–0.57 (0.04)	–0.53 (0.06)	0.31 (0.04)	–4.01 (0.36)	–0.11 (0.10)
Flowering start	–0.61 (0.10)	–0.54 (0.10)	0.38 (0.11)	–4.55 (0.89)	–0.02 (0.19)
Full flowering	–0.60 (0.10)	–0.54 (0.12)	0.38 (0.11)	–4.52 (0.97)	–0.04 (0.23)
Flowering herbs	–0.46 (0.18)	–0.42 (0.16)	0.24 (0.14)	–3.48 (1.33)	–0.04 (0.14)
Fruit maturity	–0.16 (0.11)	–0.14 (0.10)	0.04 (0.04)	–1.45 (0.86)	0.04 (0.06)
Leaf colouring	0.27 (0.08)	0.27 (0.06)	0.08 (0.04)	2.23 (0.44)	0.04 (0.09)
Leaf drop	n.a.	0.21 (0.06)	n.a.	1.75 (0.44)	n.a.

5.3 Discussion

Estimates of temperature sensitivity appear to be relatively consistent, no matter whether they are based on temporal or spatial variation, on older or recent data, and on fluctuations or trends. This particularly holds for the spring phenophases; estimates are less consistent for later phenophases. Standard errors of estimated coefficients for temperature sensitivity are generally small. It seems that temperature sensitivity is estimated precisely and reliably with the stations available in the Swiss Phenology Network.

Other studies comparing estimates of temperature sensitivity based on spatial and temporal patterns partly obtained similar estimates (*Phillimore et al.*, 2013), and partly found considerable differences (*Jochner et al.*, 2013). These contrasting results may be due to the nature of the data being analysed: The present work, similar to that of *Phillimore et al.* (2013), analyses relationships through space and through time with the same data, while *Jochner et al.* (2013) analysed two different data sets.

The finding of almost equal temperature sensitivity for the periods 1970–2000 and 1996–2012 does not imply that temperature sensitivity is constant through time: *Rutishauser et al.* (2008) calculated temperature sensitivity of spring phenology in Switzerland (an index combining several phenophases at multiple stations) for moving 30-year periods between 1750 and 2005, and found this to vary between –2 days/decade and –6 days/decade, with a decrease in temperature sensitivity from 1980 till 2005. The present result may be due to temperature sensitivity increasing again after 2005.

For summer and autumn phenophases, long-term trends suggest a higher temperature sensitivity than interannual fluctuations. *Menzel et al.* (2006) also obtained different estimates of temperature sensitivity when she related onset dates directly to monthly mean temperatures (–2.5 days/°C) and when she related mean phenological trends to mean temperature trends (ca. –6 days/°C, derived from Fig. 4 in that publication). Temperature sensitivity estimates obtained from long-term trends may be biased by any other factor that caused phenological shifts during the observation period (e.g. changes in air

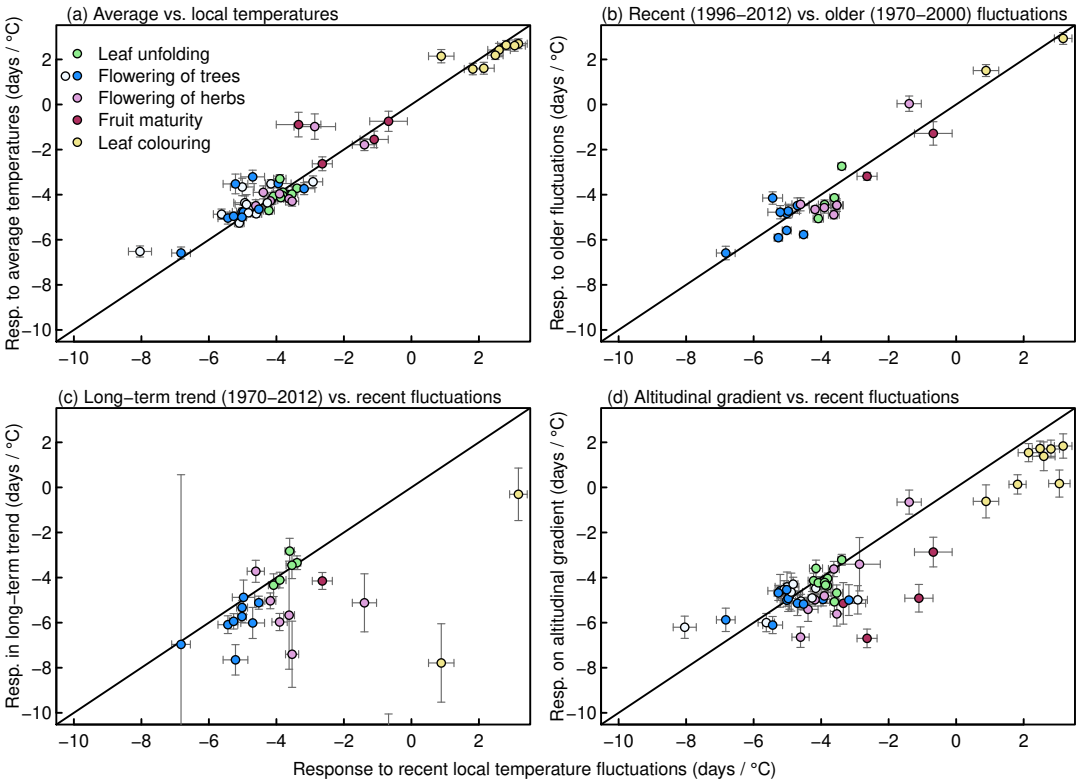


Figure 18: Comparison of temperature sensitivity coefficients derived from different types of variation (year-to-year fluctuations in local or average temperature, long-term trends, changes with altitude). Each point in the graphs represents mean sensitivity coefficients for one phenophase; error bars indicate standard errors of means based on variation among stations. The two dot colours for flowering of woody species correspond to start of flowering and full flowering, respectively. See Appendix 11 for the temperature sensitivity of particular phenophases.

quality) and are therefore less reliable than those derived from interannual fluctuations.

The increase in temperature sensitivity with altitude found for a few phenophases may partly explain why phenological trends were stronger at high altitude (Chapter 3.2). However, many phenophases with an altitude-dependent trend do not have an altitude-dependent temperature sensitivity, so that a stronger change in effective temperature is the more likely reason for the altitude-dependent trend. Previous studies have reported considerable spatial variation in temperature sensitivity (*Ziello et al.*, 2012). It has been suggested that a lack of winter chilling could reduce responses to spring temperatures in warm regions (*Primack et al.*, 2009; *Yu et al.*, 2010) or that earlier snow melt could enhance the effect of warmer temperatures on spring plant development in mountain regions (*Pellerin et al.*, 2012). The set of phenophases with altitude-dependent temperature sensitivity found here includes herbs and woody species, as well as early-spring and late-spring phenophases, so that a relationship with winter chilling requirements is not obvious. However, results do suggest that variation in temperature sensitivity among stations is not entirely random, and that it will be of particular interest to further monitor whether and how responses to temperature change with altitude. A good representation of high-altitude stations in the Swiss Phenology Network is therefore important also for future analyses of temperature sensitivity.

6 Predictive models and application to data validation

6.1 Aims and methods

Based on the results obtained so far, this last chapter presents a series of models of increasing complexity to predict phenological onset dates. The aim is to find simple models with high predictive power and to estimate prediction errors. Models are fitted separately to each of the 57 phenophases using the recent (1996–2012) data.

1. Altitudinal model: Linear model (one-way analysis of variance) with five altitudinal layers: <500, 500–799, 800–999, 1000–1199, >1200 m a.s.l. This five-layer model is currently used by MeteoSwiss to define the range of plausible values for new observations (C. Defila, unpublished data), therefore it is used here instead of a three-layer model.

$$y_{\ell i} = \mu + L_\ell + \epsilon_{\ell i} \qquad \sum_\ell L_\ell = 0 \qquad \epsilon_{\ell i} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2) \tag{2}$$

where μ is the overall mean onset date and L_ℓ is the effect of altitudinal layer ℓ. Models including altitude as numeric variable yield similar results to this one and are not presented.

2. Station model: Variance-components model with random effect of stations (α_s):

$$y_{si} = \mu + \alpha_s + \epsilon_{si} \qquad \alpha_s \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_\alpha^2) \qquad \epsilon_{si} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2) \tag{3}$$

3. Station + temperature model: Random-intercept model including the additive effects of station (α_s) and average temperatures (x_{si}):

$$y_{si} = \mu + \alpha_s + \beta \cdot x_{si} + \epsilon_{si} \tag{4}$$

where x_{si} is the average temperature at station s in year i as defined in Chapter 5.1.

The independent normal distributions assumed for α_s and ϵ_{si} in model 2 (equation 3) also hold for this and all following models.

4. Station $*$ temperature model: Random-intercept-and-slope model including the effects of station and average temperatures, as in model 3, but with a station-specific slope (β_s) for the response to temperature deviations.

$$y_{si} = \mu + \alpha_s + \beta_s \cdot x_{si} + \epsilon_{si} \qquad \beta_s \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2_\beta) \tag{5}$$

5. Station-year model: Variance-components model with the additive, independent random effects of station (α_s) and year (ζ_i).

$$y_{si} = \mu + \alpha_s + \zeta_i + \epsilon_{si} \qquad \zeta_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2_\zeta) \tag{6}$$

6. Station + temperature + nearest station : Model 3 (equation 4) is expanded by adding onset dates of the same phenophase at the nearest station ($y_{nearest}$) as predictor. Nearest stations are identified from a spatial distance matrix of stations (Euclidean distance based on x- and y-coordinates). Nearest stations are identical for all phenophases. If a phenophase is not recorded at the nearest station, the corresponding observations are excluded from model fitting. Thus, fewer observations are included in fitting this model than in fitting model 3.

$$y_{si} = \mu + \alpha_s + \beta \cdot x_{si} + \gamma \cdot y_{nearest,i} + \epsilon_{si} \tag{7}$$

7. Station + temperature + most correlated station: Model 3 (equation 4) is expanded by adding onset dates of the same phenophase at the most correlated station ($y_{correlated.s}$) as predictor. Most correlated stations are identified separately for each station and phenophase using the correlations of time series as defined in Chapter 4.1. Only correlations based on $\geq$10 pairs of observations are taken into account to increase the reliability of results. The procedure guarantees that $y_{correlated}$ has been recorded for all stations. Fig. 19 illustrates the matching for three phenophases. The most correlated stations are often far away from each other. Furthermore, the average degree of correlation between the most correlated stations differs among variables. It is generally high for leaf unfolding and flowering of woody species, and lower for leaf colouring and flowering of herbs, as illustrated in Fig. 19.

$$y_{si} = \mu + \alpha_s + \beta \cdot x_{si} + \gamma \cdot y_{correlated.s,i} + \epsilon_{si} \tag{8}$$

8. Station + temperature + most correlated variable: Model 3 (equation 4) is expanded by adding onset dates of the most correlated phenophase at the same station ($y_{correlated.p}$). Correlated phenophases are identified from a matrix of Spearman rank correlations between phenophases calculated across both stations and years. In model fitting, observations (stations and years) for which the most correlated phenophase has not been recorded are excluded, to that fewer observations are included in fitting this model than in fitting model 3.

$$y_{si} = \mu + \alpha_s + \beta \cdot x_{si} + \gamma \cdot y_{correlated.p,i} + \epsilon_{si} \tag{9}$$

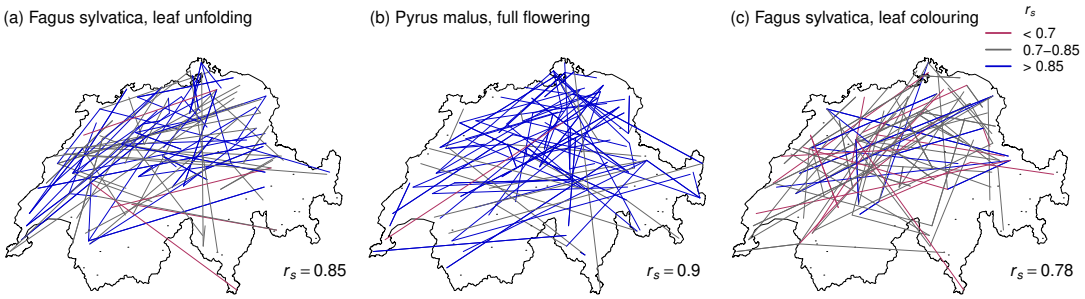


Figure 19: Spatial distance between the most correlated stations for three phenophases. Each station is linked to the station with the most correlated time series provided that $\geq$ 10 pairs of observations are available. Line colours indicate the degree of correlation (Spearman rank correlation). The mean correlation of the matched pairs of stations is given below each map.

.

9. Station + temperature + previous year: Model 3 (equation 4) is expanded by adding onset dates of the same phenophase at the same station in the previous year. Observations from 1996 (first year in the recent data set) are excluded as no previous-year data are available.

$$y_{si} = \mu + \alpha_s + \beta \cdot x_{si} + \gamma \cdot y_{s,i-1} + \epsilon_{si} \tag{10}$$

Models 10–13: Model 5 (equation 6) is expanded by including the same additional predictors as in models 6–9.

Model comparison is based on the AIC, residual standard error and mean squared prediction error from leave-one-out cross-validation. To make AIC values comparable, they are converted into AIC differences to model 2 ('stations'), which acts as reference model. AIC values can only be compared between models for identical observations. The 13 models above partly include different observations because values of the additional predictors may be missing. Therefore, separate AIC comparisons are carried out for models 1–5 (without additional predictors) and models 6–13 (with additional predictors). In the second case, the data set is first reduced to the observations with available data for all these models, then models 2, 3, 5 and 6–13 are re-fitted to these observations before calculating the AIC. Residual standard errors are calculated from all available data for each individual model. In cross-validation, some observations must be additionally excluded because they are the only representatives of a certain factor level and therefore cannot be cross-validated.

6.2 Results

The ability of linear mixed models of varying complexity to predict individual observations is compared graphically in Fig. 20 for twelve arbitrarily selected phenophases. The first series of models compares different abiotic factors as predictors, while the second and third series of models explores the possibility of using other observations as additional predictors. A simple variance components model with the random effect of individual stations (model 2) acts as reference in all model comparisons.

A simple linear model including only altitudinal layers performs worse than the 'individual-stations' model: for all phenophases, the AIC is higher (positive difference to the reference model in Fig. 20a), and the residual standard deviation is larger (Fig. 20d). Models that include annual temperature devi-

ations in addition to stations perform better than the reference (negative difference in AIC in Fig. 20a, smaller residual standard deviation in Fig. 20d). There is virtually no difference in this respect between models with a single fixed slope for the response to temperature and models with different (random) slopes for each station. Finally, models including random effects of individual years in addition to stations generally perform best among these simple models.

Including phenological observations of the same phenophases at the nearest stations into the model with temperature deviations slightly reduces the AIC for most phenophases (Fig. 20b), but the effect on the residual standard deviation is small (Fig. 20d). The same holds if the most correlated variable is included in the model. Conversely, including observations at the most correlated set of stations considerably reduces both the AIC and the residual standard deviation and thus, substantially improves model predictions (Fig. 20b, d). Including other observations as predictors into the model with effects of individual years has smaller effects on model quality (Fig. 20c, f). In particular, the 'year+station-model' is not improved further by including observations at the nearest station. The model including observations at the most correlated set of stations is again the best of the four models in Fig. 20c, f, but not better than the corresponding model based on temperature deviations (Fig. 20b, d).

Table 17 gives the mean residual standard deviations per phase of five models. The residual standard deviation of model 1 (altitude) is on average only 2.3 days smaller than the mean standard deviation of the raw data (15.3 days). Compared with this model, residual standard deviation is reduced on average by 3.1 days if individual stations are included in the model (model 2), by 1 more day if temperature deviations are included (model 3), by further 0.6 days if individual years are included (model 5), and again by further 1.5 days if the most correlated set of stations is included (model 7). With most models, residual standard deviation is smallest for leaf unfolding and flowering of trees and shrubs, and largest for fruit maturity (Table 17).

The prediction error (error made when applying a model to new observations) is estimated by cross-validation (Table 18). Results for the same 12 phenophases as before indicate that the prediction error is nearly identical to the residual standard deviation with model 1 (altitude), and 0.3–0.5 days higher with models 2–7, which have more parameters. The difference is small, and prediction errors strongly correlate with residual standard deviations (not shown), hence the relative performance of the models is identical with both error measures.

Table 17: Standard deviation of residuals from five models, calculated for each variable and then averaged for each of the phenological stages. Models are: (1) five altitudinal layers, (2) individual stations, (3) individual stations+temperature deviations, (5) individual stations+individual years, (7) individual stations+temperature deviations+observations at the most correlated set of stations

Stage	model 1	model 2	model 3	model 5	model 7
Leaf unfolding	10.12	8.05	6.94	6.42	5.13
Flowering start	12.87	9.91	8.47	7.59	6.13
Full flowering	12.58	9.70	8.20	7.43	6.06
Fruit maturity	18.61	13.47	13.32	12.49	10.24
Leaf colouring	13.26	10.23	9.90	9.58	7.95
Leaf drop	12.58	9.83	9.63	9.14	7.36
Flowering of herbs	14.75	10.87	9.93	9.29	7.61
All stages	13.03	9.96	9.00	8.35	6.80

Table 18: Prediction error of five models, determined through cross-validation (root mean squared prediction error) for 12 phenophases. The last two lines are the mean prediction error of the 12 phenophases and, for comparison, the mean residual standard deviation (as in Table 17) of the same 12 phenophases. Models are: (1) five altitudinal layers, (2) individual stations, (3) individual stations+temperature deviations, (5) individual stations+individual years, (7) individual stations+temperature deviations+observations at the most correlated set of stations

Species	stage	model 1	model 2	model 3	model 5	model 7
Fagus sylvatica	LU	8.02	6.96	5.76	5.13	4.17
Tilia platyphyllos	LU	9.51	8.22	6.56	6.13	4.91
Larix decidua	LU	11.18	9.39	8.40	7.45	6.10
Corylus avellana	FF	19.63	17.55	13.48	12.28	9.46
Sambucus racemosa	FF	16.02	10.45	9.16	8.98	7.38
Pyrus malus	FF	9.14	8.16	6.27	5.63	4.58
Tussilago farfara	FH	15.00	13.50	11.98	11.09	8.97
Cardamine pratensis	FH	12.07	9.76	8.54	7.87	6.13
Epilobium angustifolium	FH	25.51	16.91	16.80	16.66	13.13
Fagus sylvatica	LC	13.03	10.10	9.61	9.25	7.67
Tilia platyphyllos	LC	13.59	11.52	11.15	10.86	8.46
Larix decidua	LC	11.84	9.26	8.91	8.69	6.87
Mean prediction error		13.71	10.98	9.72	9.17	7.32
Mean residual standard deviation		13.69	10.47	9.25	8.66	7.00

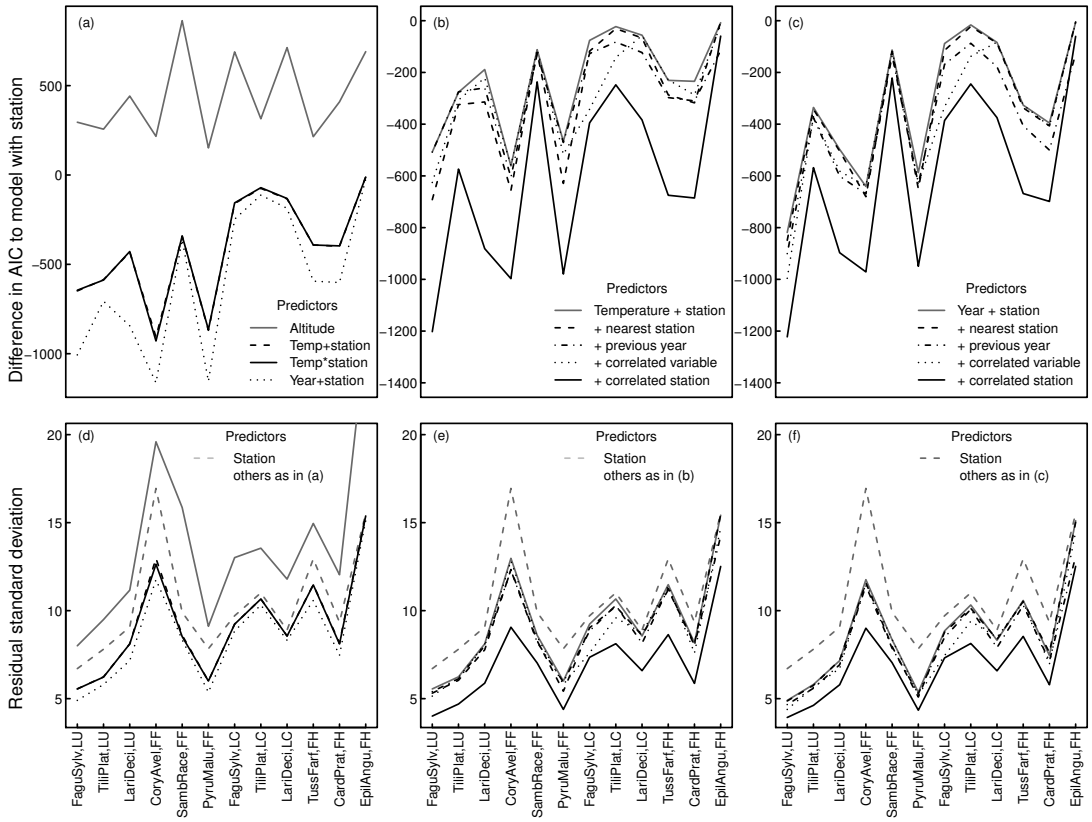


Figure 20: Comparison of models including different predictors for 12 of the variables. (a, b, c) Comparison based on the AIC. For each phenophase, the difference in AIC to model 2: $y_{si} = \mu + \alpha_s + \epsilon_{si}$ (s = station, i = year) is given. Negative differences indicate better predictive models. (d, e, f) Comparison based on residual standard deviation. Smaller values indicate better model fit. Predictors considered are: *station*: random effect of individual stations, *altitude*: five altitudinal classes, *temperature*: deviation of country-wide average temperatures from the 1996–2012 mean for the 2 months preceding and including the median date of the event in three groups of stations, *year*: random effect of individual years, *nearest station*: onset date of the same phenophase at the set of stations with smallest spatial distance to the actual stations, *correlated phenophase*: onset date of the most correlated correlated phenophase at the same set of stations, *correlated station*: onset date of the same phenophase at the set of stations most correlated with the focal stations for this particular phenophase. In (a) 'Temp+station' means additive effects, i.e. same response to temperature deviations at all stations (model 3), while 'Temp*station' means interactive effects, i.e. a different slope for the response to temperature deviations at each station (model 4). In (b) and (c), all compared models include the additive effects of temperature deviations or years, plus one series of values as additional predictor, corresponding to models 3 and 6–9 in (b), and to models 5 and 10–13 in (c).

.

6.3 Discussion

The comparison of models with different predictors confirms that variability in the data is not optimally accounted for by considering only the effects of temperature, i.e. altitude and temperature deviations. Model fit is considerably improved by accounting for the effects of individual stations and years. Various recent analyses using hierarchical models showed the importance of including local effects as source of variability in the data (*Primack et al.*, 2009; *Ibáñez et al.*, 2010). While those studies focused on the uncertainty introduced by that variability when trying to forecast future phenological changes (*Ibáñez et al.*, 2010), local effects can also contribute to making data more predictable as long as the set of stations remains identical.

The limited predictive value of temperature deviations in the model comparison could be partly due to the use of average rather than local temperatures, a relatively short time window (2 months), and the fact that this window was defined by the median date per group of stations, and not for each individual station. These choices were made to simplify computations (avoiding variable- and station-specific adjustments) in view of a possible application in data quality checking. However, using other time windows does not greatly change the results. Re-running model 3 with different time windows indicates that the residual standard deviation is reduced on average by 0.01 days if time windows are defined for each individual stations, by 0.09 days if 3-month windows are used, and by 0.24 days if local temperatures are used (detailed results not shown). Furthermore, model 4 (with station-specific responses to temperature) does not perform better than model 3, in contrast to results from other networks covering a larger geographic and climatic gradients (*Ibáñez et al.*, 2010; *Primack et al.*, 2009).

Effects of individual years predict the data better than temperature deviations, suggesting that additional time-varying factors influence phenology, such as precipitation (*Studer et al.*, 2005; *Jentsch et al.*, 2009; *Wielgolaski*, 2001) or sunshine duration. Given that data on these factors are available from the SwissMetNet network, it would be worth checking their possible contribution to data prediction. For practical purposes (data quality checking during online data entry), model 5 has the disadvantage that the current year's effect is unknown until all data from this year are available. This may lead to a preference for model 3 despite poorer fit.

Attempts to improve predictions by including other observations in the models were mostly ineffective, as could be expected from the inconsistent correlation patterns reported above. The models used here includes single slope coefficients, e.g. for the relationship between observations at the focal station and those at the nearest station (model 6); the relationship is therefore assumed to be identical at all stations. In a preliminary analysis, predictive models fitted separately to individual stations were partly improved by including the nearest station, but such a multitude of models would not be applicable in practice. For similar reasons, although the inclusion of station-specific temporal autocorrelation in `gls` models slightly increases the precision of estimated coefficients for interannual variation (Table 8), the simpler approach of including previous-year data as covariate (model 8) does not improve the predictions.

At first sight, including observations at the most correlated station (model 7) seems to substantially improve our ability to predict individual observations. But this model includes many 'hidden parameters'

due to the selection of the most correlated station out of 138 potential ones for each individual station and phenophase. Such 'hidden parameters' are expected to cause an overfitting that is not detected by penalized measures of model fit such as the AIC. This is why cross-validation is performed to assess prediction error. Contrary to expectation, cross-validation error is not much higher than residual standard deviation, and smaller than for the other models, suggesting that model 7 is the best predictive model. However, when correlations between pairs of stations in the period 1996–2012 (the basis for model 7) are compared with those determined for the same pairs of stations in the period 1970–1995, there is virtually no relationship between the two correlation matrices ($r < 0.1$ in all cases), and the identity of the most correlated stations differs in 99% of the cases. Thus, while model 7 performs well in predicting observations in the period for which it has been fitted, it is unlikely to perform well in the future. Such a model would regularly have to be updated, making it unpractical. Thus, model 3 (with stations and temperature deviations) and model 5 (with stations and years) are probably the most suitable models for data quality checking.

6.4 Application to data validation

The online data entry interface *PhaenoNet* and since 2018 *Phenotool* includes an automatic plausibility check. When new data are entered online by an observer, they are automatically compared to expected values. If the deviation exceeds a limit, a warning is issued, inviting the observer to check his entry. This automatic system is aimed at filtering out mistakes, such as recording the wrong event or incorrectly reconstructing the date of observation if data are recorded retrospectively. To be effective, warnings should be generated with high probability when incorrect data are entered, so that mistakes can be corrected (sensitivity), but only rarely when correct data are entered, to avoid unnecessary annoyance (specificity).

Currently, initial data checking is based on means and standard deviations of all observations of the respective variable in each of five altitudinal layers. This corresponds to the fitted values and residual standard deviation of model 1 above, except that model 1 assumes the standard deviation to be identical in all layers. Warnings are issued if newly entered data deviate by more than 2 standard deviations from the expected value (mean of that layer). We shall now examine how the sensitivity and specificity could be optimized by using different models to define expected values and standard deviations.

Assuming that residuals are normally distributed, and that warnings are generated if new data deviate by more than k standard deviations from the expected values, the probability of a 'false' warning (upon entry of correct data) is $2 \cdot \Phi(-k)$, where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. With a threshold k of 2, 2.5, 3 or 3.5, the frequency of false warnings is 4.55%, 1.24%, 0.27% and 0.05%, respectively, no matter how the expected values and standard deviations are defined. Thus, specificity depends only on the choice of the threshold k. The larger k, the higher the specificity.

Conversely, sensitivity depends both on k (larger values of k reduce the sensitivity) and on the residual standard deviation. The more precisely a model predicts correct data, the smaller is the residual standard deviation, and the larger is the probability that an error of a certain size will generate a warning. Indeed, for additive observation errors with a certain size (e.g. being 1 week or 1 month wrong when trying to reconstruct the date of an event), erroneous observations follow a normal distribution

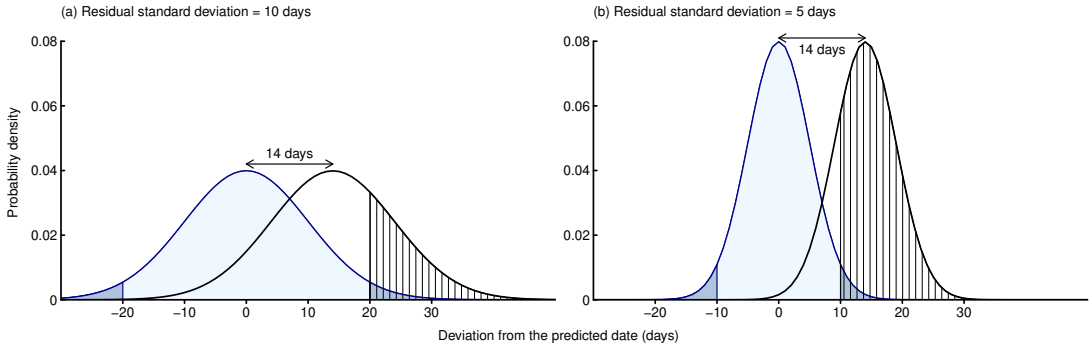


Figure 21: Relationship between the precision of predictions and sensitivity of error detection. The reference distribution, i.e. the theoretical distribution of observations around their expected value, is represented by the blue curve. This is taken to be a normal distribution with standard deviation of (a) 10 days and (b) 5 days (as two examples). The black curve represents the distribution of observations modified by an additive error of two weeks. The limits $\pm$ 2·sd define the range of values for which data entry generates a warning. Blue areas represent the probability of false warnings (1-specificity), while hatched areas represent the probability of warnings for erroneous data (sensitivity). In (a) the large standard deviation leads to a small relative error size (δ), so that erroneous data often fail to generate a warning. In (b), the smaller standard deviation leads to a larger relative error size, and most erroneous data generate a warning. Note that specificity is identical in both cases.

with the same standard deviation but a shifted mean compared to the reference distribution (Fig. 21). The size of the error relative to the standard deviation of the reference distribution (relative error size, δ) determines the probability of error detection. The larger δ, the more clearly separated are the distributions of correct and erroneous data (Fig. 21). The probability that an erroneous observation generates a warning (sensitivity) is $\Phi(-k-\delta)+\Phi(-k+\delta)$. As illustrated by Fig. 21, the same absolute error generates a warning with much higher probability if the reference distribution has a small variability (Fig. 21a) than if it has a large variability (Fig. 21b) because the relative error size δ is larger in the first case.

Fig. 22 further illustrates how the probability of error detection (sensitivity) depends on the frequency of false warnings as determined by the threshold k with various relative error sizes δ. The threshold k has a strong influence on sensitivity at small relative error size, while the effect becomes marginal at large relative error size (Fig. 22a). As a simple thought experiment, Fig. 22b further assumes that the probability of erroneous data being actually checked and corrected upon a warning decreases linearly with increasing frequency of false warnings. This leads to an optimal frequency of warnings (optimal threshold k), which decreases with increasing relative error size (Fig. 22b). Thus, large relative errors make it possible to achieve a high probability of errors being corrected with a low frequency of annoying false warnings.

Table 19 gives the sensitivities for various combinations of values for k, absolute error size and residual standard deviation of the model defining expected values. A 1-week error rarely generates a warning regardless of the threshold k and even with a precise model (residual sd of 5). A 2-week error often generates a warning with a residual standard deviation of 5 ($\delta = 3$) but not otherwise. A 1-month error almost always generates a warning with a residual error of 5 or 7.5, and a 2-month error almost alway does so ($\delta > 4$). As noted before, for sufficiently large values of δ, increasing k from 2 to 2.5 increases specificity more than it decreases sensitivity. For example, with a residual sd of 5 days, a 2-week error generates a warning with 79% probability if k = 2 and with 62% probability if $k = 2.5$. At the same time, the frequency of false warnings decreases more than threefold from 4.55% to 1.24%.

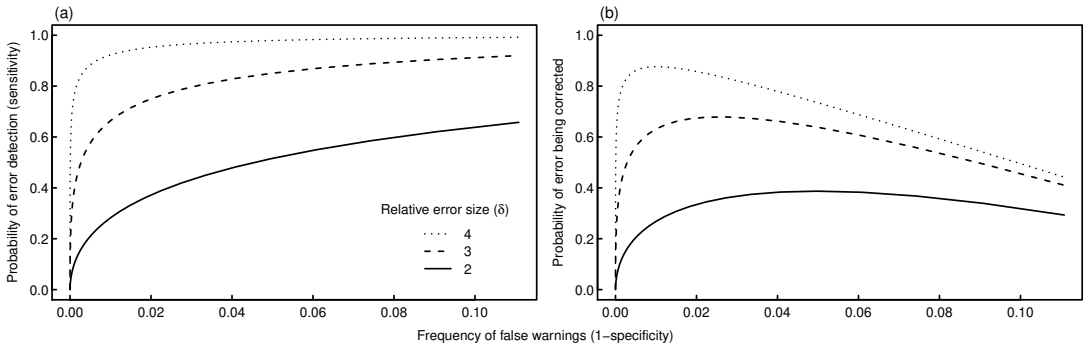


Figure 22: Relationships between the relative frequency of false warnings (for correct data) and (a) the probability of erroneous data generating a warning or (b) the probability of erroneous data being actually checked and corrected, assuming that the probability of an observer critically checking his entry after a warning decreases from 1 to 0.5 as the frequency of false warnings increases from 0 to 0.1. Note that this assumption is purely hypothetical.

Table 19: Probability of detecting errors of different size with different warning thresholds (k = 2, 2.5 or 3) based on reference distributions with different standard deviations (sd).

Error	sd = 5 days			sd = 7.5 days			sd = 10 days			sd = 15 days		
Size	k = 2	2.5	3	k = 2	2.5	3	k = 2	2.5	3	k = 2	2.5	3
1 week	0.27	0.14	0.05	0.14	0.06	0.02	0.10	0.04	0.01	0.07	0.02	0.01
2 weeks	0.79	0.62	0.42	0.45	0.26	0.13	0.27	0.14	0.05	0.14	0.06	0.02
1 month	1.00	1.00	1.00	0.98	0.93	0.84	0.84	0.69	0.50	0.50	0.31	0.16
2 months	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.98	0.93	0.84

Thus, with reference to the model comparison above (Tables 17 and 18), the use of a more precise yet still simple model, such as model 3, in combination with $k = 2.5$, would make it possible to improve the sensitivity and specificity of the automatic data quality check. However, errors of 1–2 weeks would still often remain unnoticed. Using simulated data, *Schaber and Badeck* (2002) reached similar conclusions. These authors additionally showed that robust estimation procedures lead to more frequent error detection than linear models because estimates of error variation are less influenced by outliers (*Schaber and Badeck*, 2002). The present analysis was carried out after outlier exclusion, therefore the use of linear models seemed sufficient.

7 Conclusions

- Phenological variation across Switzerland is determined by altitude, large-scale spatial trends (e.g. climatic regions) and unspecified local factors (including variation among individual plants and observation error), whereas small-scale spatial dependence is marginal for most phenophases. Mean onset dates, interannual variation and long-term trends of some phenophases vary with altitude and climatic regions.
- Altitudinal gradients and spatial trends are significant, yet a large fraction of phenological variation is local and unrelated to the spatial location of stations. Hence, the precision of overall parameter estimates depends more on the total number of stations than on their distribution over the country. Additional stations will contribute largely independent additional information even if they are only 10 km away from existing ones.
- Since altitudinal gradients and spatial trends can be of ecological interest, a good representation of the different altitudinal layers and climatic regions remains desirable. Because relationships of phenology with altitude can vary regionally, an ideal distribution of stations would include several stations from each combination of region and altitudinal layers. Currently, high elevations in the Jura and southern Switzerland and low elevations in the central Alps are less represented.
- The availability of phenological stations over an altitudinal range of 1.7 km (200–1900 m a.s.l.) within a small geographic area is a particular asset of the Swiss Phenology Network. It provides a unique opportunity to study the role of altitude for phenological trends with minimal confounding effects such as changes in latitude and continentality.
- The number of stations currently included in the Swiss Phenology Network is sufficient for precise estimates of mean onset dates of each phenophase, of long-term trends and of responses to temperature for the entire country and for three altitudinal layers. For a precise analysis of regional differences, more stations would be needed.
- There are no groups of stations with similar patterns of interannual variation for all phenophases, i.e. none of the phenological stations provides largely redundant information to the network.
- Neighbouring stations are hardly more similar to each other than any other pair of stations. Thus, phenological observations at neighbouring stations do not contribute to the prediction or verification of new phenological data. The use of other phenophases or previous-year data as covariates also fails to improve the predictive power of models. However, a model that includes the effects of individual stations and years would predict observations more precisely than a model based only on altitudinal groups, leading to more effective error detection.

List of Figures

List of Tables

References

Beaubien, E. G., and A. Hamann (2011), Plant phenology networks of citizen scientists: recommendations from two decades of experience in Canada, *International Journal of Biometeorology*, *55*, 833–841.

Begert, M. (2008), Die Repräsentativität der Stationen im Swiss National Basic Climatological Network (Swiss NBCN). Arbeitsberichte der MeteoSchweiz Nr. 217, 40 pp.

Both, C., M. van Asch, R. G. Bijlsma, A. B. van den Burg, and M. E. Visser (2009), Climate change and unequal phenological changes across four trophic levels: constraints or adaptations?, *Journal of Animal Ecology*, *78*, 73–83.

Brügger, R., and A. Vassella (2003), *Pflanzen im Wandel der Jahreszeiten. Anleitung für phänologische Beobachtungen - Les plantes au cours des saisons. Guide pour observations phénologiques*, Geographica Bernensia, Bern.

Chmielewski, F.-M. (2013), Phenology in agriculture and horticulture, in *Phenology: An Integrative Environmental Science*, edited by M. D. Schwartz, pp. 539–561, Springer, Dordrecht.

Chmielewski, F.-M., and T. Rötzer (2001), Response of tree phenology to climate change across Europe, *Agricultural and Forest Meteorology*, *108*, 101–112.

Cleland, E. E., I. Chuine, A. Menzel, H. A. Mooney, and M. D. Schwartz (2007), Shifting plant phenology in response to global change, *Trends in Ecology and Evolution*, *22*, 357–365.

Cleland, E. E., J. M. Allen, T. M. Crimmins, J. A. Dunne, S. Pau, S. E. Travers, E. S. Zavaleta, and E. M. Wolkovich (2012), Phenological tracking enables positive species responses to climate change, *Ecology*, *93*, 1765–1771.

Defila, C., and B. Clot (2001), Phytophenological trends in Switzerland, *International Journal of Biometeorology*, *45*, 203–207.

DeGaetano, A. T. (2001), Spatial grouping of United States climate stations using a hybrid clustering approach, *International Journal of Climatology*, *21*, 791–807.

Dierenbach, J., F.-W. Badeck, and J. Schaber (2013), The plant phenological online database (PPODB): an online database for long-term phenological data, *International Journal of Biometeorology*, *57*, 805–812.

Dittmar, C., and W. Elling (2006), Phenological phases of common beech (*Fagus sylvatica* L.) and their dependence on region and altitude in Southern Germany, *European Journal of Forest Research*, *125*, 181–188.

Dose, V., and A. Menzel (2004), Bayesian analysis of climate change impacts on phenology, *Global Change Biology*, *10*, 259–272.

Estrella, N., T. H. Sparks, and A. Menzel (2009), Effects of temperature, phase type and timing, location, and human density on plant phenological responses in Europe, *Climate Research*, *39*, 235–248.

Fitter, A., and R. S. R. Fitter (2002), Rapid changes in flowering time in British plants, *Science*, *296*, 1689–1691.

Furrer, R., and S. R. Sain (2009), Spatial model fitting for large datasets with applications to climate and microarray problems, *Statistics and Computing*, *19*, 113–128.

Gehrig, R. (2012), Die Repräsentativität der Pollenmessstationen des Schweizer Pollenmessnetzes. Arbeitsberichte der MeteoSchweiz, 237, 76 pp.

Güsewell, S. (2014), *Phenological responses to changing temperatures: representativeness and precision of results from the Swiss Phenological Network*, Master Thesis in Biostatistics, University of Zurich.

Güsewell, S., R. Furrer, R. Gehrig, and B. Pietragalla (2017), Changes in temperature sensitivity of spring phenology with recent climate warming in Switzerland are related to shifts of the preseason, *Global Change Biology*, *23*, 5189–5202.

Ibáñez, I., R. B. Primack, A. J. Miller-Rushing, E. Ellwood, H. Higuchi, S. D. Lee, H. Kobori, and J. A. Silander (2010), Forecasting phenology under global warming, *Philosophical Transactions of the Royal Society B*, *365*, 3247–3260.

IPCC (2012), Summary for policymakers, in *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change*, edited by C. B. Field and V. Barros, pp. 3–21, Cambridge University Press, Cambridge, New York.

Jentsch, A., J. Kreyling, J. Boettcher-Treschkow, and C. Beierkuhnlein (2009), Beyond gradual warming: extreme weather events alter flower phenology of European grassland and heath species, *Global Change Biology*, *15*, 837–849.

Jochner, S., A. Caffarra, and A. Menzel (2013), Can spatial data substitute temporal data in phenological modelling? A survey using birch flowering, *Tree Physiology*, *13*, 1256–1268.

Koch, E. (2010), Global framework for data collection: data bases, data availability, future networks, online databases, in *Phenological Research: Methods for Environmental and Climate Change Analysis*, edited by I. J. Hudson and M. R. Keatley, pp. 23–61, Springer, Dordrecht.

Linkosalo, T. (2000), Mutual regularity of spring phenology of some boreal tree species: predicting with other species and phenological models, *Canadian Journal of Forest Research*, *30*, 667–673.

Linkosalo, T., R. Häkkinen, and P. Hari (1996), Improving the reliability of a combined phenological time series by analyzing observation quality, *Tree Physiology*, *16*, 661–664.

Mächler, M., P. Rousseeuw, A. Struyf, M. Hubert, and K. Hornik (2013), *cluster: Cluster analysis basics and extensions*, R package version 1.14.4.

Maignan, F., F. M. Bréon, E. Vermote, P. Ciais, and N. Viovy (2008), Mild winter and spring 2007 over western Europe led to a widespread early vegetation onset, *Geophysical Research Letters*, *35*, L02,404.

Menzel, A. (2013), Europe, *Phenology: An Integrative Environmental Science*, pp. 53–65.

Menzel, A., N. Estrella, and P. Fabian (2001), Spatial and temporal variability of the phenological seasons in Germany from 1951 to 1996, *Global Change Biology*, *7*, 657–666.

Menzel, A., H. Seifert, and N. Estrella (2011), Effects of recent warm and cold spells on European plant phenology, *International Journal of Biometeorology*, *55*, 921–932.

Menzel, A., et al. (2006), European phenological response to climate change matches the warming pattern, *Global Change Biology*, *12*, 1969—1976.

Miller-Rushing, A. J., T. T. Hoye, D. W. Inouye, and E. Post (2010), The effects of phenological mismatches on demography, *Philosophical Transactions of the Royal Society B*, *365*, 3177–3186.

Morisette, J. T., et al. (2009), Tracking the rhythm of the seasons in the face of global change: phenological research in the 21st century, *Frontiers in Ecology and the Environment*, *7*, 253–260.

Nychka, D., R. Furrer, and S. Sain (2013), *fields: Tools for spatial data*, R package version 6.9.1.

Parmesan, C. (2007), Influences of species, latitudes and methodologies on estimates of phenological response to global warming, *Global Change Biology*, *13*, 1860–1872.

Pellerin, M., A. Delestrade, G. Mathieu, O. Rigault, and N. G. Yoccoz (2012), Spring tree phenology in the Alps: effects of air temperature, altitude and local topography, *European Journal of Forest Research*, *131*, 1957–1965.

Phillimore, A. B., K. Proios, N. O'Mahony, R. Bernard, A. M. Lord, S. Atkinson, and R. J. Smithers (2013), Inferring local processes from macro-scale phenological pattern: a comparison of two methods, *Journal of Ecology*, *101*, 774–783.

Pinheiro, J. C., and D. M. Bates (2000), *Mixed-Effects Models in S and S-PLUS*, Springer, New York.

Primack, R. B., I. Ibáñez, H. Higuchi, S. D. Lee, A. J. Miller-Rushing, A. M. Wilson, and J. Silander, J A (2009), Spatial and interspecific variability in phenological responses to warming temperatures, *Biological Conservation*, *142*, 2569–2577.

Primault, B. (1955), Cinq ans d'observations phénologiques systématiques en suisse, *Annalen der Schweizerischen Meteorologischen Zentralanstalt*, *92*, 7/4–7/5.

Roetzer, T., M. Wittenzeller, H. Haeckel, and J. Nekovar (2000), Phenology in central Europe differences and trends of spring phenophases in urban and rural areas, *International Journal of Biometeorology*, *44*, 60–66.

Rutishauser, T., J. Luterbacher, C. Defila, D. Frank, and H. Wanner (2008), Swiss spring plant phenology 2007: Extremes, a multi-century perspective, and changes in temperature sensitivity, *Geophysical Research Letters*, *35*, L05,703.

Schaber, J., and F.-W. Badeck (2002), Evaluation of methods for the combination of phenological time series and outlier detection, *Tree Physiology*, *22*, 973–982.

Schaber, J., F. Badeck, D. Doktor, and W. von Bloh (2010), Combining messy phenological time series, in *Phenological research: methods for environmental and climate change analysis*, edited by I. L. Hudson and M. R. Keatley, pp. 147–158, Springer, Dordrecht.

Schleip, C., T. H. Sparks, N. Estrella, and A. Menzel (2009), Spatial variation in onset dates and trends in phenology across Europe, *Climate Research*, *39*, 249–260.

Schwartz, M. D. (2013), *Phenology: an Integrative Environmental Science*, Springer, Dordrecht.

Schwartz, M. D., E. G. Beaubien, T. M. Crimmins, and J. F. Weltzin (2013), North America, in *Phenology: An Integrative Environmental Science*, edited by M. D. Schwartz, pp. 67–89, Springer, Dordrecht.

Sparks, T. H., B. Jaroszewicz, M. Krawczyk, and P. Tryjanowski (2009), Advancing phenology in Europe's last lowland primeval forest: non-linear temperature response, *Climate Research*, *39*, 221–226.

Studer, S., C. Appenzeller, and C. Defila (2005), Inter-annual variability and decadal trends in alpine spring phenology: a multivariate analysis approach, *Climatic Change*, *73*, 395–414.

Studer, S., R. Stöckli, C. Appenzeller, and P. L. Vidale (2007), A comparative study of satellite and ground-based phenology, *International Journal of Biometeorology*, *51*, 405–414.

Vitasse, Y., and D. Basler (2013), What role for photoperiod in the bud burst phenology of European beech, *European Journal of Forest Research*, *132*, 1–8.

Vitasse, Y., G. Hoch, C. F. Randin, A. Lenz, C. Kollas, J. F. Scheepens, and C. Körner (2013), Elevational adaptation and plasticity in seedling phenology of temperate deciduous tree species, *Oecologia*, *171*, 663–678.

Wielgolaski, F. E. (2001), Phenological modifications in plants by various edaphic factors, *International Journal of Biometeorology*, *45*, 196–202.

Wüthrich, C., S. Scherrer, M. Begert, M. Croci-Maspoli, C. Marty, G. Seiz, N. Foppa, T. Konzelmann, and C. Appenzeller (2010), Die langen Schneemessreihen der Schweiz Eine basisklimatologische Netzanalyse und Bestimmung besonders wertvoller Stationen mit Messbeginn vor 1961. Arbeitsberichte der MeteoSchweiz Nr. 233, 33 pp.

Yu, H., E. Luedeling, and J. Xu (2010), Winter and spring warming result in delayed spring phenology on the Tibetan Plateau, *Proceedings of the National Academy of Sciences*, *107*, 22,151–22,156.

Ziello, C., N. Estrella, M. Kostova, E. Koch, and A. Menzel (2009), Influence of altitude on phenology of selected plant species in the Alpine region (1971–2000), *Climate Research*, *39*, 227–234.

Ziello, C., A. Böck, N. Estrella, D. Ankerst, and A. Menzel (2012), First flowering of wind-pollinated species with the greatest phenological advances in Europe, *Ecography*, *35*, 1017–1023.

Acknowledgments

The phenological data analysed here have been collected over decades by a large number of dedicated volunteers. Their invaluable contribution to the Swiss Phenology Network is gratefully acknowledged.

We further thank Michael Begert and Christian Sigg (MeteoSwiss) for providing the phenology and temperature database extracts, and Bernard Clot for reviewing a draft of this report.

Reinhard Furrer acknowledges support of the University of Zurich Research Priority Program (URPP) on ”Global Change and Biodiversity”.

Appendix

Appendix 1: List of stations with station number, abbreviated name, coordinates, altitude, overall period of inclusion in the phenological network, number of years with records from 1970 and from 1996, and years without records.

Appendix 2: List of phenophases with species name and phenological stage, abbreviated species name and stage, first year recorded ('start'), and number of stations (n) at which the phenophase is recorded.

Appendix 3: Distributions of correlations between species for individual phenological stages based on years 1996–2012.

Appendix 4: Regression of mean onset dates (1996–2012) against altitude. Each dot represents one station, with colours indicating climatic regions. Phenophases are sorted by phenological stage to visualize differences among stages. Regression lines are given with 95% confidence bands and regression statistics (slope b in days/100 m), residual error in days.

Appendix 5: Distribution of deviations from the overall altitudinal trend among phenophases for each station, based on mean onset dates for 1996–2012.

Appendix 6: Long-term year-to-year changes in phenology and linear trend (1970–2012) for each phenophase at all stations. Regression lines (with 95% confidence intervals) are derived from mixed models including random effects of stations. Regression slopes $\pm$ se are given. Negative slopes indicate a trend towards earlier onset dates. A negative trend can be regarded as statistically significant ($p < 0.05$) if mean$+1.96\cdot$se< 0.

Appendix 7: Relationship between long-term trends (1970-2012) and altitude for all phenophases. Each dot represents one station. Regression lines are blue if the relationship is statistically significant ($p < 0.05$) and grey otherwise.

Appendix 8: Clustering of stations: dendrograms resulting from hierarchical clustering with complete linkage based on correlations between the time series of stations (1996–2012) for four selected phenophases per phenological stage. Symbols below dendrograms indicate the altitude and region of stations.

Appendix 9A: Distribution of correlations between the time series (1996–2012) of individual stations and the mean time series of all stations as a measure of each station's representativeness. Correlations were calculated separately for each phenophase; distributions of these correlations (a) over all phenophases and (b–e) over 8–11 species per phenological stage are represented by boxplots. Stations are ordered by increasing altitude from left to right, and colours indicate climatic regions.

Appendix 9B: Distribution of correlations between the time series (1996–2012) of individual stations and the mean time series of each climatic region. Correlations were calculated separately for each phenophase; distributions of these correlations (a) over all phenophases and (b–e) over 8–11 species per phenological stage are represented by boxplots. Stations are ordered by climatic regions.

Appendix 10: Altitudinal gradients in temperature, based on norm values 1981–2010 for monthly mean temperatures at 89 climatic stations. Regression lines and slope coefficients are given in red for all stations, and in blue separately for stations above and below 700 m a.s.l.

Appendix 11: Temperature sensitivity of each phenophase (days/℃), i.e. difference in onset date associated with a 1℃ higher temperature, derived from four types of spatial or temporal variation in phenology and temperature: altitudinal gradients, year-to-year fluctuations 1996–2012, year-to-year fluctuations 1970–2012, and linear trends 1970–2012.

Appendix 12: Distribution of temperature sensitivity coefficients derived from linear trends between 1970 and 2012. (a) Distribution of phenological trends (slopes of linear regression against years), (b) distribution of the associated temperature trends, and (c) distribution of sensitivity coefficients calculated as the ratio between phenological trends and temperature trends. All coefficients were determined for each phenophase and station; boxes represents distributions over stations for each phenophase.

Appendix 1: List of stations with station number, abbreviated name, coordinates, altitude, overall period of inclusion in the phenological network, number of years with records from 1970 and from 1996, and years without records.

station	abbr	xcoord	ycoord	alt	period	from1970	from1996	missing
56	DST	708000	173000	1200	1956-2012	42	16	1964, 1998
201	VAS	734000	164000	1250	1956-2012	43	17	
338	ANR	752000	163000	985	1953-2012	43	17	1958-1969
361	TUS	753150	175100	700	1956-2012	43	17	
442	DAD	783000	187000	1560	1951-2012	43	17	1957-1959, 1965, 1969
521	LEZ	760000	176000	1500	1970-2012	43	17	
588	DOM	754000	189000	580	1970-2012	43	17	
603	ARS	770000	183000	1900	1992-2007	16	12	
639	CHR	760340	192940	640	1962-2012	39	16	1963-1964, 1984, 1987-1988, 2005
642	ZIZ	762000	200000	600	1971-2012	42	17	
695	JEN	773000	200000	800	1970-2012	42	16	2009
702	SCR	771000	204000	700	1951-2012	42	16	1955, 1957-1969, 2009
710	FAN	769000	206000	910	1970-2007	35	9	1996-1998
712	GRU	768000	206000	650	1970-2007	38	12	
719	SEW	766910	203920	960	1951-2012	43	17	1952, 1956-1969
798	SGS	752000	213000	480	1956-2012	43	17	
817	AZM	755000	216000	480	1980-2012	33	17	
862	BUH	754000	225000	450	1999-2012	14	14	
876	EBO	757640	231875	580	1973-2012	38	16	1975, 1996
981	HED	750110	256960	800	1975-2012	37	16	2000
1011	ROR	755000	260000	450	1977-2012	36	17	
1058	RON	745800	269900	405	1978-2008	30	12	2006
1239	DIH	698000	283000	410	1971-2012	42	17	
1279	MEH	688000	291000	540	1959-2012	43	17	
1309	NHA	688370	281920	435	1974-2012	39	17	
1355	RAF	683000	275000	515	1952-2012	42	17	1977
1429	HLU	676480	283560	430	1979-2012	34	17	
1444	OST	679000	279000	410	1977-2012	36	17	
1462	LBU	647000	268000	330	1970-2012	43	17	
1519	MOI	756620	209200	305	1956-2006	36	11	1961, 1966, 1968, 1973
1668	LIT	622000	259000	350	1951-2012	43	17	
1722	MTI	596000	137000	530	1963-2012	43	17	1964
1761	BEL	579350	234970	930	1956-2012	40	15	1988, 1998, 2001
1781	RGS	583170	247930	865	1978-2012	33	15	1996-1997
1823	BAI	602604	247824	486	2010-2010	1	1	
1850	GLI	612150	254560	330	2004-2012	9	9	
1892	DOR	613000	259000	300	1970-2012	43	17	
1925	THE	609000	261000	310	2003-2012	10	10	
1941	BAB	610820	266440	315	1966-2012	41	16	1984, 2009
2018	WIH	745000	230000	1100	1951-2012	43	17	
2082	WAT	726000	239000	625	1951-2012	43	17	
2136	WIL	722000	258000	600	1991-2012	22	17	
2201	FWI	731570	253070	630	1952-2012	31	17	1966-1967, 1969, 1980-1991
2283	APL	749400	244230	775	1956-2012	43	17	1969
2302	PST	744230	249660	825	1979-2012	34	17	
2348	SGW	747000	225000	655	2004-2012	9	9	
2401	HBW	728000	270000	510	1992-2011	20	16	
2417	RAW	721000	277000	595	1992-2012	21	17	
2558	LAW	725000	264000	570	1991-2010	20	15	
2601	FFE	708660	267430	405	1992-2012	21	17	
2629	GUD	704000	266000	460	1987-2010	23	14	1998
2658	GUN	700000	276500	460	2004-2006	3	3	
2711	BUA	709000	247000	640	1957-2012	35	10	1962-1968, 1993, 2001-2007
2778	WTH	697000	262000	580	1951-2012	21	17	1953-1955, 1957-1965, 1968, 1970-1991
2855	WEK	704000	242000	555	1992-2012	20	16	1998
3075	MUG	735000	219000	500	1951-2012	43	17	
3125	LIN	719000	197000	650	1970-2012	43	17	
3182	ELP	732070	198220	1000	1956-2012	43	17	
3255	NAF	724000	217000	440	1970-2012	38	17	1987-1991
3442	WAL	713850	237250	620	1955-2010	36	13	1956-1961, 1963-1972, 1997-1998
3541	WAD	694000	231000	480	1956-2012	43	17	1960-1968
3561	HOR	686540	235560	450	1966-2012	42	16	1999
3629	ZWI	687000	246000	620	1965-2012	42	16	2006
3701	ZHP	685125	248090	555	1955-2012	43	17	
3761	HYB	705000	208000	1500	1992-2012	20	17	1993
3799	EIS	699400	220940	910	1958-2011	37	16	1960-1967, 1969-1974
3838	SIG	686200	230450	480	1951-2012	34	10	1962-1968, 1985-1986, 2001-2007
3865	ZAL	681000	245000	530	1952-2012	41	15	1962-1968, 2005-2006
3903	BID	676000	244000	500	1984-2012	29	17	
4121	SLN	694000	182000	510	1991-2012	22	17	
4139	ALD	693000	192000	470	1975-2012	34	14	1983, 1996-1998
4559	SNN	662000	194000	500	1954-2012	43	17	1964
4589	LZE	666920	210500	600	1975-2012	38	17	
4649	ENB	647950	204230	765	1958-2012	43	17	
4661	WOL	649000	212000	580	1992-2012	21	17	
4775	EDB	686000	226000	750	1971-2012	35	10	2000-2006
5051	GAM	669580	176050	1205	1956-2012	42	17	1973
5078	MEG	658000	175000	850	1956-2012	23	17	1972-1991
5166	WEG	637000	161500	1300	1993-2012	18	15	2005, 2007
5194	UNT	631800	170500	565	2007-2012	6	6	
5231	KAN	617430	149740	1175	1956-2012	43	17	
5269	ADB	609400	148975	1350	1956-2012	43	17	1958-1959, 1962-1964
5289	RFR	616000	159000	1090	1992-2012	21	17	
5351	ZWS	595000	156000	965	1958-2012	43	17	1959-1969
5469	OLG	623050	184800	950	1978-2009	32	14	

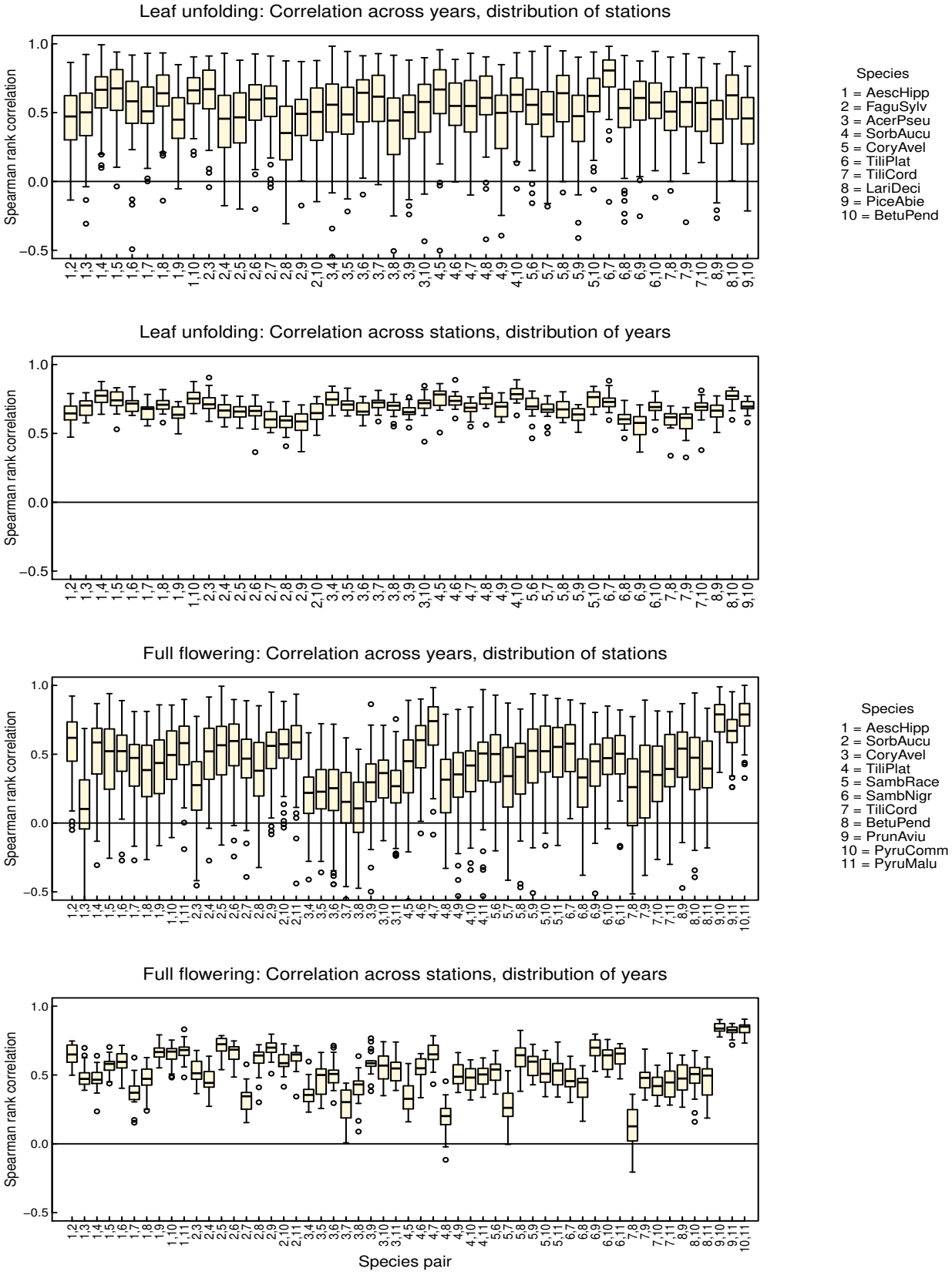
Appendix 1 (cont.): List of stations with information about years with available records.

station	abbr	xcoord	ycoord	alt	period	from1970	from1996	missing
5495	HOF	610000	174000	720	1991-2012	19	14	2009-2011
5508	MSN	609205	191960	540	2008-2009	2	2	
5529	WOR	609000	197000	600	1986-2012	27	17	
5588	GST	588260	146165	1045	2004-2012	9	9	
5649	VSA	580390	166420	1050	1956-2012	43	17	1958
5742	POS	574000	179000	680	1951-2012	41	15	1952-1958, 1960-1968, 1997-1998
5871	MDN	551000	169000	500	1966-2012	41	15	1998-1999
6052	VOR	519000	174000	850	1951-2010	41	15	1953-1957
6069	ABT	527540	178770	660	1956-2012	42	17	1986
6092	OBO	531000	175000	485	1953-2012	35	17	1961-1977
6173	ELA	555000	188000	450	1989-2012	24	17	
6231	COE	539000	197000	750	1954-2012	43	17	
6238	LPM	546000	206000	1120	1951-2012	43	17	
6299	BVI	536130	203780	1050	1979-2012	34	17	
6326	CER	559000	212000	800	1951-2012	43	17	1953
6343	BOY	554000	200000	450	2004-2012	9	9	
6351	CHT	565000	211000	1150	1994-2012	19	17	
6352	ENS	567585	211830	820	1951-2012	43	17	1953, 1964-1965
6371	BNN	585620	220130	490	1951-2012	32	17	1964, 1981-1991
6392	MTS	566200	223500	1200	1993-2012	20	17	
6402	ORV	583000	223000	700	1951-2012	42	16	1968, 2000
6469	EHT	637620	194290	910	1956-2012	43	17	
6532	JET	605200	210900	525	2007-2012	5	5	2009
6533	SEL	597800	210800	560	2007-2012	6	6	
6539	JEG	605001	215050	520	1984-2012	27	15	1996, 2001
6579	OES	613250	219525	485	1960-2012	40	14	2004-2006
6589	HZB	620000	226000	450	1969-2012	30	17	1979-1991
6592	WYA	628000	234000	450	1970-2012	42	17	1971
6599	WYS	629000	215000	850	1970-2012	42	17	1971
6605	GDW	634000	224000	610	1963-2009	40	14	1966
6671	ZON	638000	238000	440	1951-1999	30	4	1957, 1967-1968
6765	SCW	642330	246200	450	1970-2012	43	17	
6815	WIB	644540	236690	650	1952-2012	43	17	
6820	KOE	644250	242700	430	2007-2012	6	6	
6820	HOC	664500	224300	490	2007-2012	6	6	
6905	SEO	655000	244000	550	1952-2012	43	17	
6919	MRI	667630	235320	550	1975-2012	35	14	1996-1998
6952	VIL	654000	258000	415	1970-2012	43	17	
6972	DOT	661050	269000	350	1975-2012	31	10	2000-2006
6993	OHD	667350	259400	490	1975-2012	38	17	
7069	FIS	653000	139000	1100	1951-2012	40	15	1962-1964, 1989, 2004-2005
7251	VIP	635000	127000	650	1992-2006	14	11	1993
7330	SLU	612000	119000	1650	1970-2012	43	17	
7573	LET	583000	115000	480	1976-2012	37	17	
7642	LID	580000	93000	1350	1951-2012	40	14	1957, 2004-2006
7715	TRT	565980	99790	1300	1951-2012	43	17	
7801	GON	571250	124890	1100	1956-2012	43	17	
7821	PLB	572530	121990	1100	1957-2012	39	14	1960-1965, 1975, 2000-2001, 2010
7877	MOG	554540	122100	1380	1975-2012	38	17	
7941	DIL	577710	133080	1200	1958-2008	35	13	1964-1965, 1980-1981, 1983-1984
7957	SEY	571310	134960	1265	1978-2012	33	15	1996-1997
7964	LEY	568000	132000	1250	1967-2012	43	17	1969
8029	BLO	559000	146000	600	1988-2010	23	15	
8043	CDO	553000	147000	655	1983-2012	28	15	1997-1998
8261	LOG	509980	150390	900	1959-2012	41	15	1996-1997
8291	CHI	507000	139000	435	1965-2011	42	16	
8338	VES	502000	126000	440	1952-2012	43	17	
8452	CAR	490000	115000	400	1954-2012	43	17	1955-1957
8539	LOL	548000	211000	1020	1956-2012	42	16	2006
9091	MOT	716000	145000	500	1988-2012	23	16	1993, 2009
9131	SBD	734490	147180	1625	1956-2005	36	10	1958, 1960, 1962
9279	VIR	708950	111790	210	1975-2012	38	17	
9309	PSO	692000	139000	750	1957-2012	43	17	
9318	CEC	689370	130000	430	1966-2012	41	15	2002-2003
9352	AUR	698000	121000	350	1953-2012	36	10	1954-1955, 1957, 2002-2008
9353	VEG	690000	120000	1100	1956-2012	42	17	1990
9402	LOM	704160	114350	370	1991-2012	22	17	
9403	LCN	705000	114000	200	1966-2012	43	17	
9441	RAC	718700	81000	353	2008-2012	5	5	
9449	SGO	725000	79000	670	1987-2012	25	16	2008
9578	SID	647720	116140	1475	1952-2012	42	16	1996
9709	BPI	806330	127300	800	1956-2012	40	17	1984-1986
9743	CAS	771000	140000	1460	1970-2012	43	17	
9759	SPP	766000	135000	1000	1970-2012	42	17	1985
9778	BOD	763000	134000	825	1970-2012	43	17	
9829	SMO	784300	152420	1800	1956-2012	43	17	1959-1968
9851	POA	789000	152000	1780	1970-2012	43	17	
9858	ZUO	793000	164000	1750	1970-2012	42	17	1989
9931	SCL	818000	187000	1240	1971-2012	42	17	
9932	SEN	821000	189000	1440	1970-2009	40	14	
9959	MTN	830580	197100	1050	1976-2012	37	17	
9981	STM	828000	165000	1390	1998-2012	15	15	

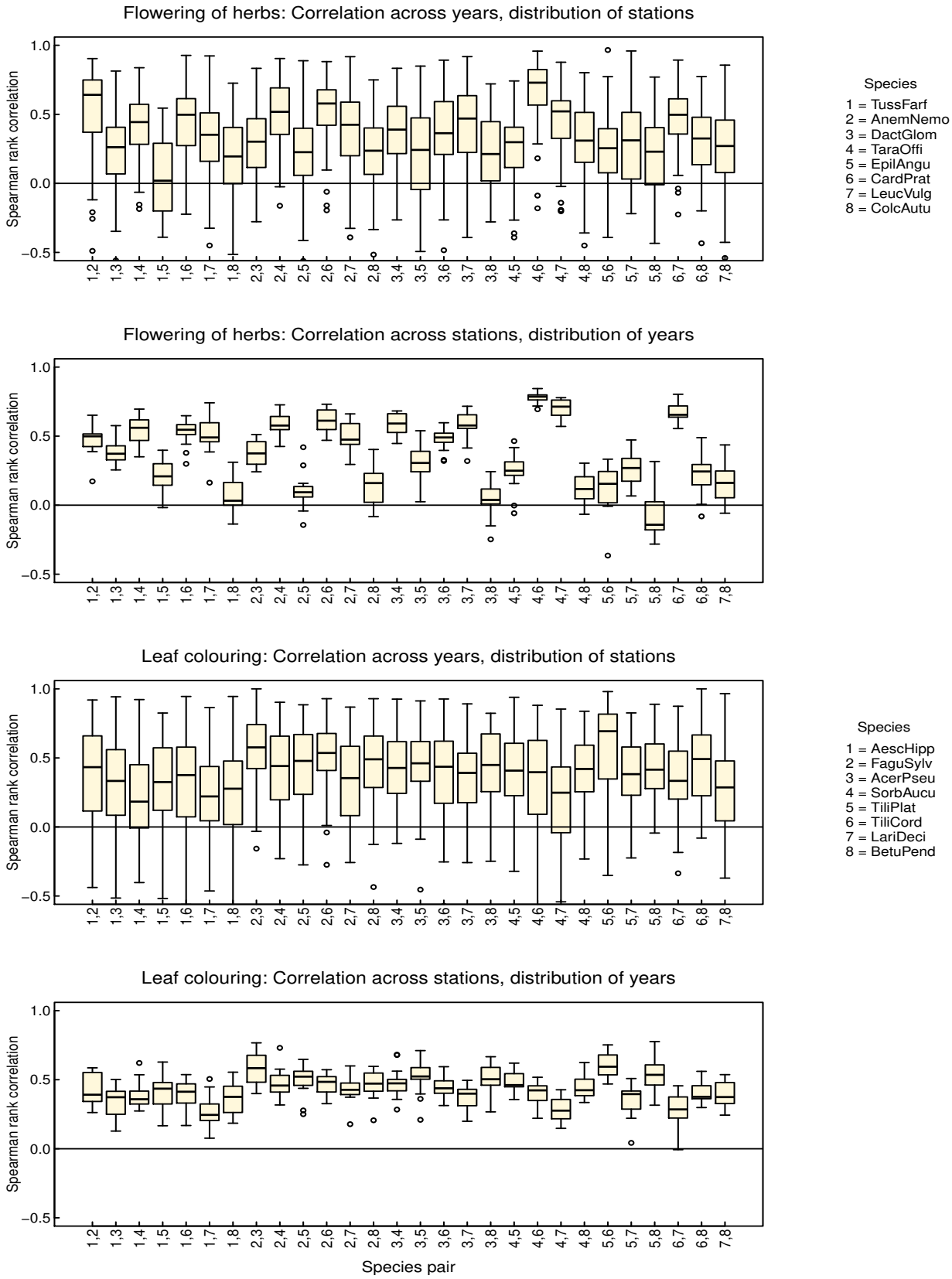
Appendix 2: List of phenophases with species name and phenological stage, abbreviated species name and stage, first year recorded ('start'), and number of stations (*n*) at which the phenophase is recorded.

variable	species name	phenological stage	species	stage	start	*n*
kjd00100	Aesculus hippocastanum	leaf unfolding	AescHipp	LeafUnfolding	1951	141
kjd00200	Aesculus hippocastanum	beginning of flowering	AescHipp	FlowerStart	1996	133
kjd00300	Aesculus hippocastanum	full flowering	AescHipp	FlowerFull	1951	141
kjd00400	Aesculus hippocastanum	leaf discoloration	AescHipp	LeafColor	1951	141
kjd00500	Aesculus hippocastanum	leaf drop	AescHipp	LeafDrop	1951	140
kjd00600	Fagus sylvatica	leaf unfolding	FaguSylv	LeafUnfolding	1951	151
kjd00700	Fagus sylvatica	leaf discoloration	FaguSylv	LeafColor	1951	147
kjd00800	Fagus sylvatica	leaf drop	FaguSylv	LeafDrop	1951	146
kjd00900	Acer pseudoplatanus	leaf unfolding	AcerPseu	LeafUnfolding	1996	146
kjd01000	Acer pseudoplatanus	leaf discoloration	AcerPseu	LeafColor	1996	144
kjd01100	Sorbus aucuparia	leaf unfolding	SorbAucu	LeafUnfolding	1996	155
kjd01200	Sorbus aucuparia	beginning of flowering	SorbAucu	FlowerStart	1996	149
kjd01300	Sorbus aucuparia	full flowering	SorbAucu	FlowerFull	1996	154
kjd01400	Sorbus aucuparia	fruit maturity	SorbAucu	Fruit	1953	159
kjd01500	Sorbus aucuparia	leaf discoloration	SorbAucu	LeafColor	1996	151
kjd01600	Sorbus aucuparia	leaf drop	SorbAucu	LeafDrop	1996	151
kjd01700	Corylus avellana	leaf unfolding	CoryAvel	LeafUnfolding	1951	158
kjd01800	Corylus avellana	beginning of flowering	CoryAvel	FlowerStart	1996	155
kjd01900	Corylus avellana	full flowering	CoryAvel	FlowerFull	1952	156
kjd02000	Tilia platyphyllos	leaf unfolding	TiliPlat	LeafUnfolding	1996	140
kjd02100	Tilia platyphyllos	beginning of flowering	TiliPlat	FlowerStart	1996	136
kjd02200	Tilia platyphyllos	full flowering	TiliPlat	FlowerFull	1951	147
kjd02300	Tilia platyphyllos	leaf discoloration	TiliPlat	LeafColor	1995	135
kjd02400	Sambucus racemosa	beginning of flowering	SambRace	FlowerStart	1996	140
kjd02500	Sambucus racemosa	full flowering	SambRace	FlowerFull	1951	148
kjd02600	Sambucus racemosa	fruit maturity	SambRace	Fruit	1996	131
kjd02700	Sambucus nigra	beginning of flowering	SambNigr	FlowerStart	1996	155
kjd02800	Sambucus nigra	full flowering	SambNigr	FlowerFull	1951	158
kjd02900	Sambucus nigra	fruit maturity	SambNigr	Fruit	1996	151
kjd03000	Tilia cordata	leaf unfolding	TiliCord	LeafUnfolding	1996	142
kjd03100	Tilia cordata	beginning of flowering	TiliCord	FlowerStart	1996	134
kjd03200	Tilia cordata	full flowering	TiliCord	FlowerFull	1951	146
kjd03300	Tilia cordata	leaf discoloration	TiliCord	LeafColor	1996	138
kjd03400	Larix decidua	needle emergence	LariDeci	LeafUnfolding	1951	163
kjd03500	Larix decidua	needle discoloration	LariDeci	LeafColor	1996	161
kjd03600	Larix decidua	needle drop	LariDeci	LeafDrop	1996	157
kjd03700	Picea abies	needle emergence	PiceAbie	LeafUnfolding	1951	163
kjd03800	Robinia pseudoacacia	leaf unfolding	RobiPseu	LeafUnfolding	1996	89
kjd03900	Robinia pseudoacacia	beginning of flowering	RobiPseu	FlowerStart	1996	84
kjd04000	Robinia pseudoacacia	full flowering	RobiPseu	FlowerFull	1996	89
kjd04100	Robinia pseudoacacia	leaf drop	RobiPseu	LeafDrop	1996	85
kjd04200	Betula pendula	leaf unfolding	BetuPend	LeafUnfolding	1996	156
kjd04300	Betula pendula	beginning of flowering	BetuPend	FlowerStart	1996	142
kjd04400	Betula pendula	full flowering	BetuPend	FlowerFull	1996	142
kjd04500	Betula pendula	leaf discoloration	BetuPend	LeafColor	1996	149
kjd04600	Betula pendula	leaf drop	BetuPend	LeafDrop	1996	150
kjd04700	Castanea sativa	leaf unfolding	CastSati	LeafUnfolding	1996	67
kjd04800	Castanea sativa	beginning of flowering	CastSati	FlowerStart	1996	54
kjd04900	Castanea sativa	full flowering	CastSati	FlowerFull	1996	56
kjd05000	Castanea sativa	fruit maturity	CastSati	Fruit	1996	54
kjd05100	Castanea sativa	leaf discoloration	CastSati	LeafColor	1996	59
kjd05200	Castanea sativa	leaf drop	CastSati	LeafDrop	1996	59
kjd05300	Tussilago farfara	full flowering	TussFarf	FlowerHerb	1951	158
kjd05400	Anemone nemorosa	full flowering	AnemNemo	FlowerHerb	1951	151
kjd05500	Dactylis glomerata	full flowering	DactGlom	FlowerHerb	1996	151
kjd05600	Taraxacum officinale	full flowering	TaraOffi	FlowerHerb	1951	166
kjd05700	Epilobium angustifolium	full flowering	EpilAngu	FlowerHerb	1996	122
kjd05800	Cardamine pratensis	full flowering	CardPrat	FlowerHerb	1951	154
kjd05900	Leucanthemum vulgare	full flowering	LeucVulg	FlowerHerb	1951	165
kjd06000	Colchicum autumnale	full flowering	ColcAutu	FlowerHerb	1953	150
kjd06100	Prunus avium	beginning of flowering	PrunAviu	FlowerStart	1996	153
kjd06200	Prunus avium	full flowering	PrunAviu	FlowerFull	1951	154
kjd06300	Pyrus communis	beginning of flowering	PyruComm	FlowerStart	1996	144
kjd06400	Pyrus communis	full flowering	PyruComm	FlowerFull	1951	145
kjd06500	Pyrus malus	beginning of flowering	PyruMalu	FlowerStart	1996	150
kjd06600	Pyrus malus	full flowering	PyruMalu	FlowerFull	1951	154
kjd06700	Vitis vinifera	full flowering	VitiVini	FlowerFull	1951	79
kjd06800	Vitis vinifera	vintage	VitiVini	Fruit	1951	81
kjd06900	Hay harvest	start	HayHarve	Fruit	1951	164

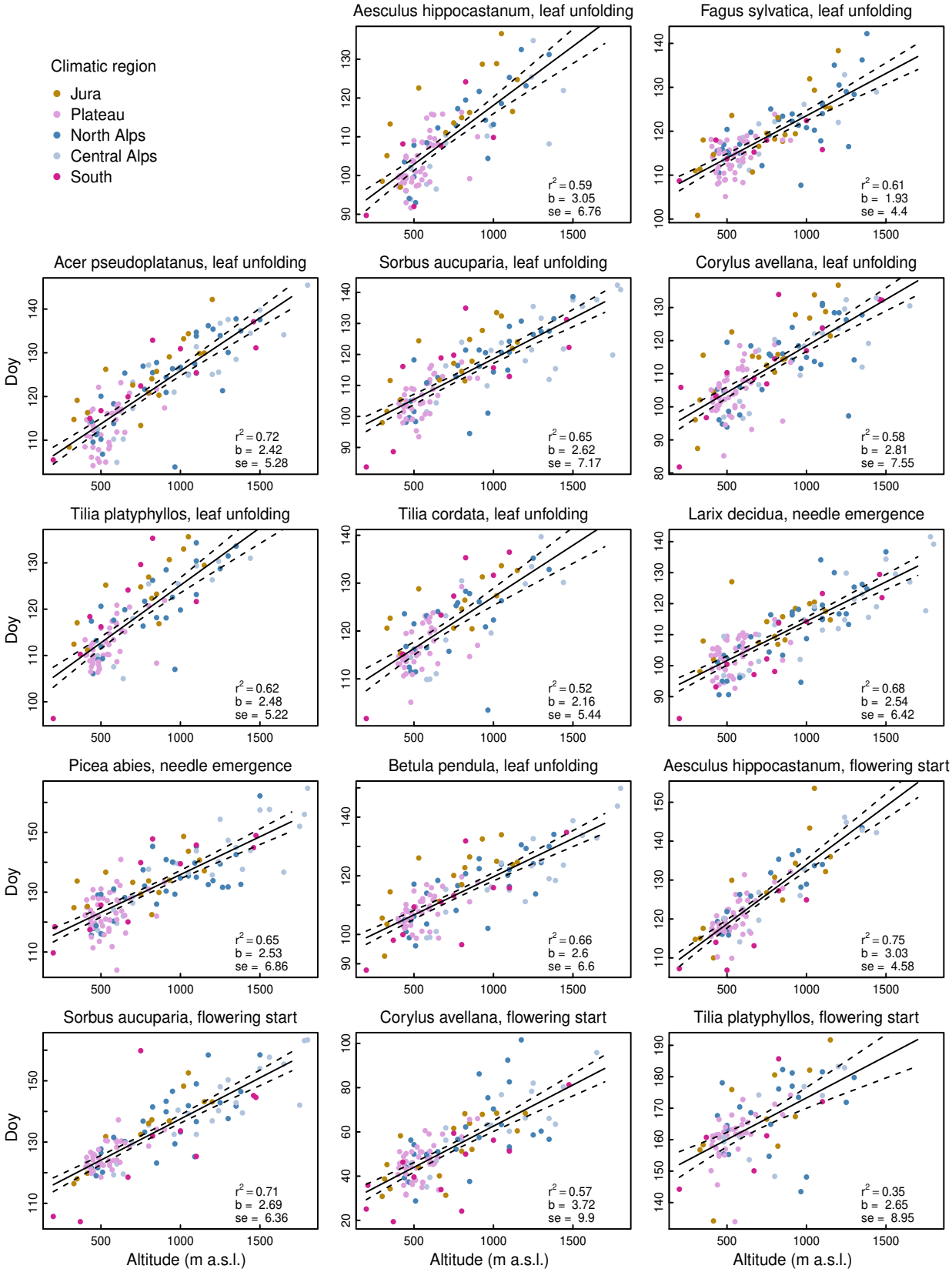
Appendix 3: Distributions of correlations between species for individual phenological stages based on years 1996–2012.



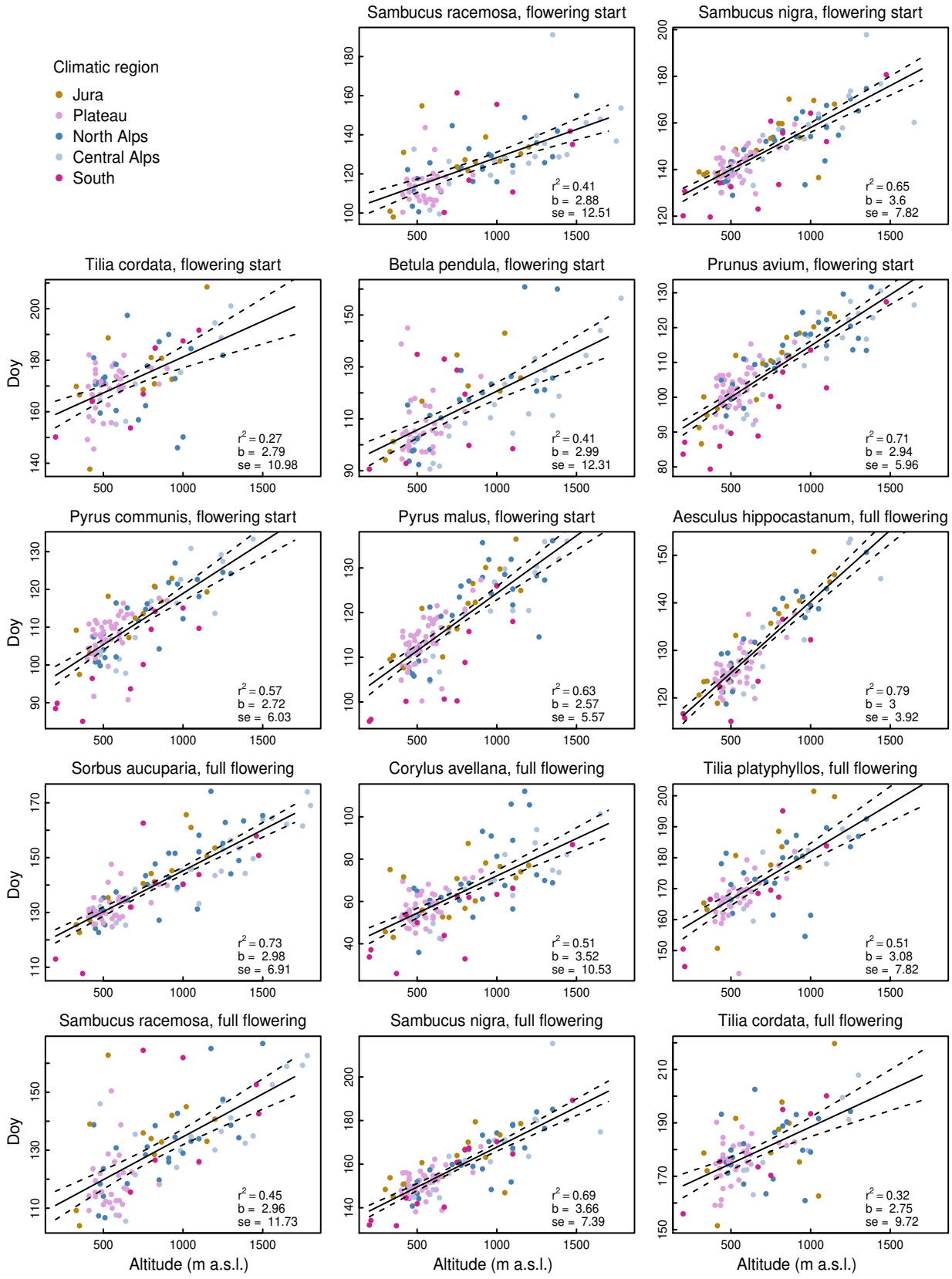
Appendix 3 (cont.): Distributions of correlations between species for individual phenological stages.



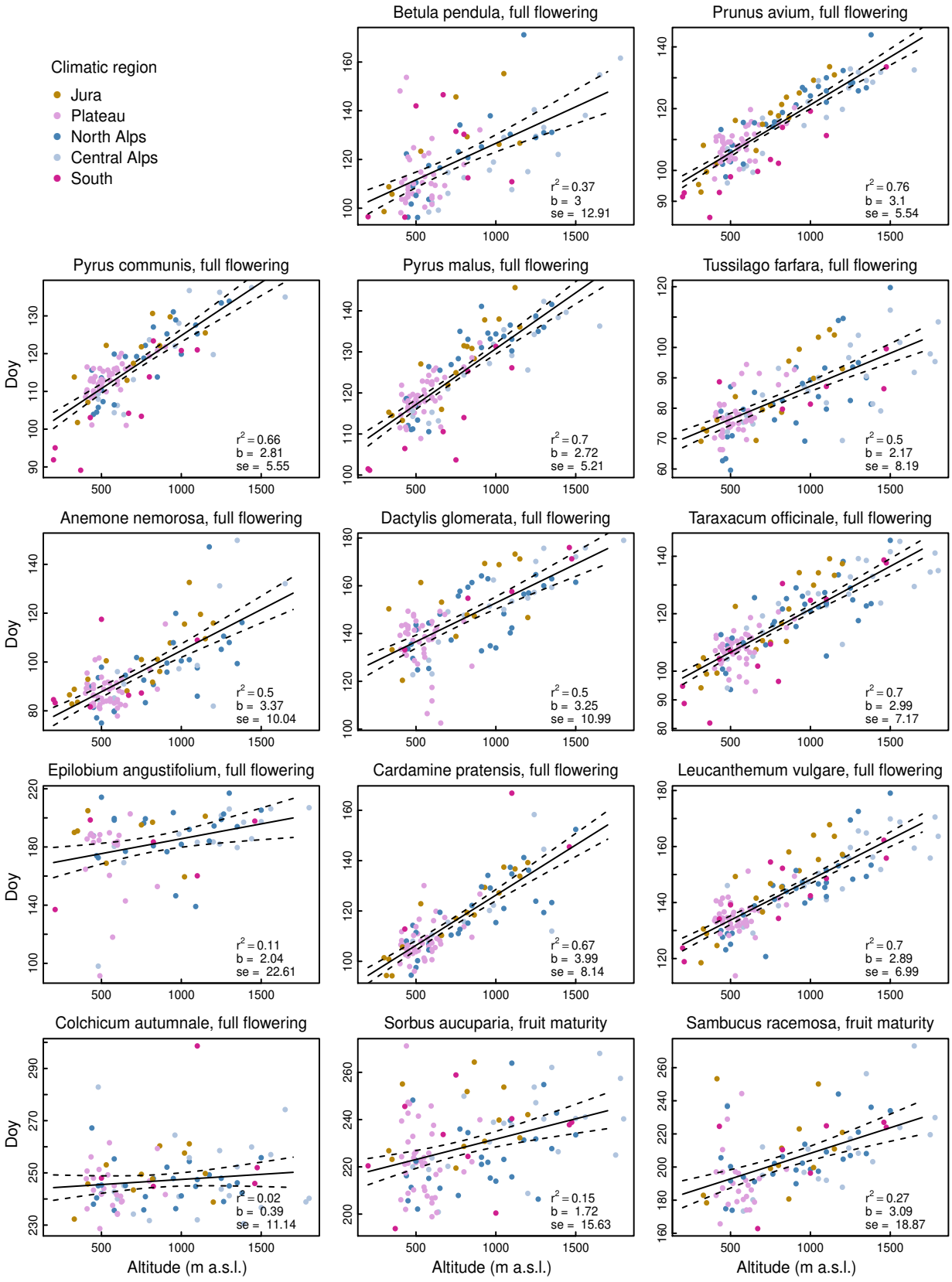
Appendix 4: Regression of mean onset dates (1996–2012) against altitude. Each dot represents one station, with colours indicating climatic regions. Phenophases are sorted by phenological stage to visualize differences among stages. Regression lines are given with 95% confidence bands and regression statistics (slope *b* in days/100 m), residual standard error (se) in days.



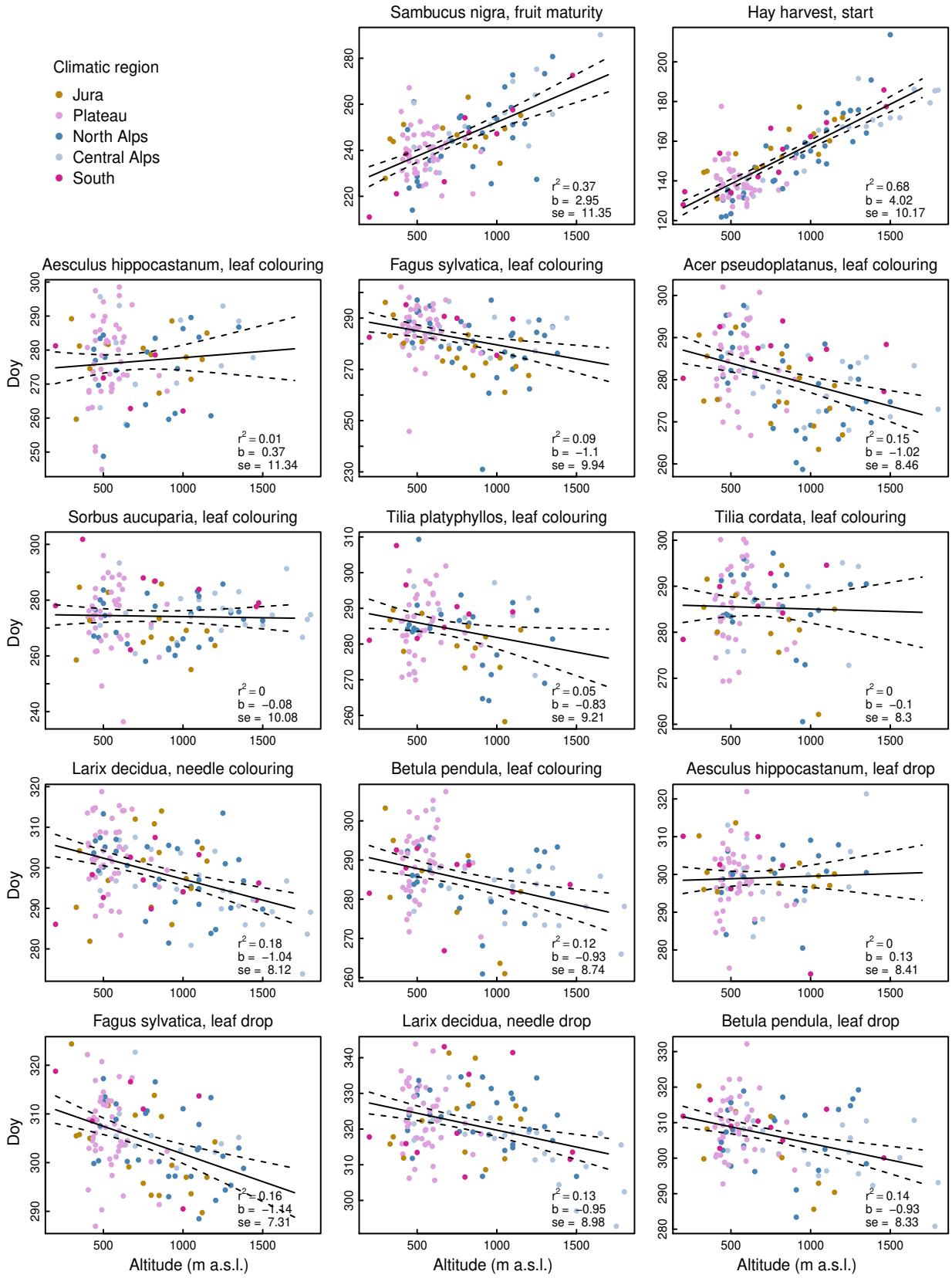
Appendix 4 (cont.): Regression of mean onset dates (1996–2012) against altitude.



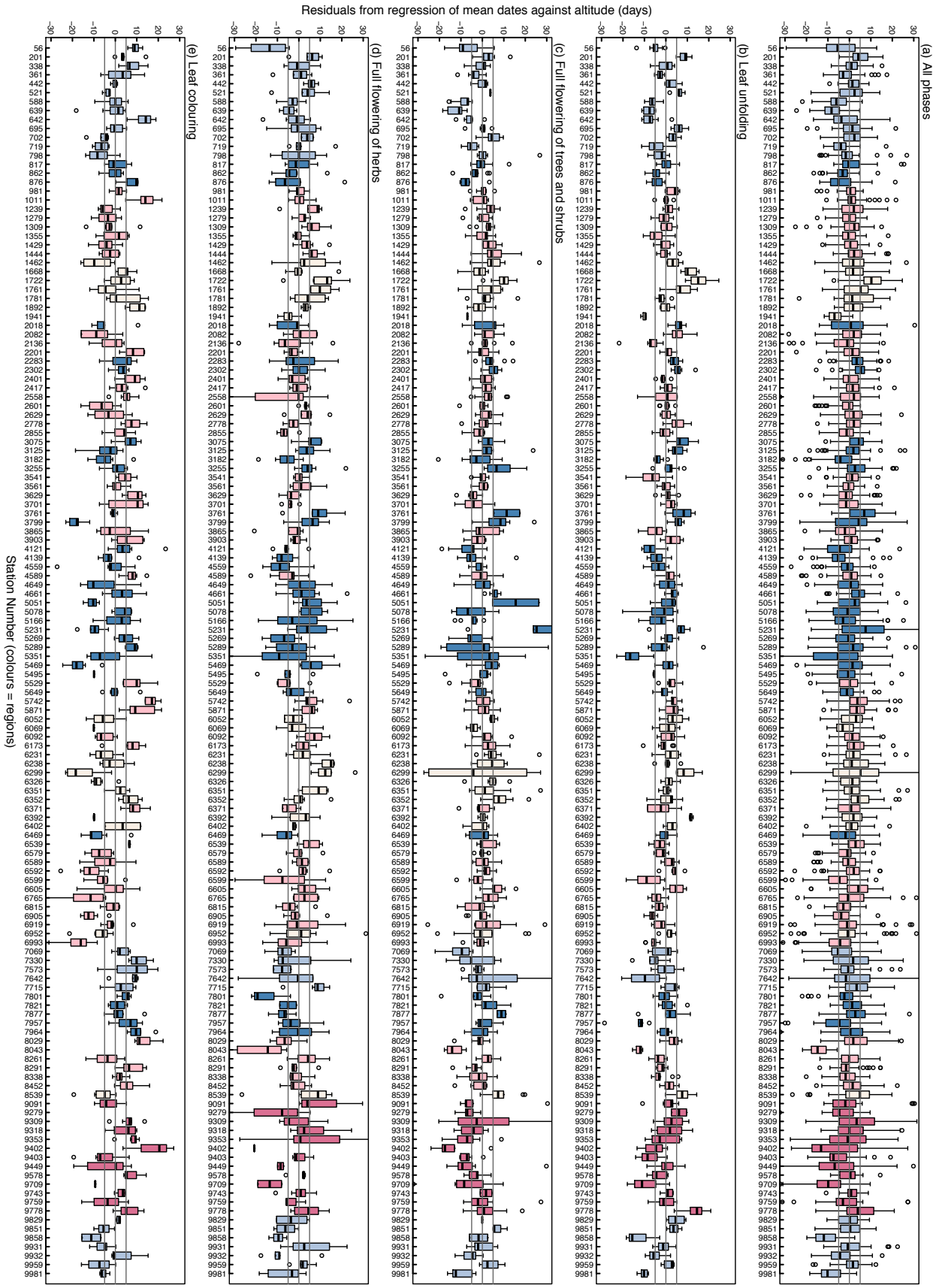
Appendix 4 (cont.): Regression of mean onset dates (1996–2012) against altitude.



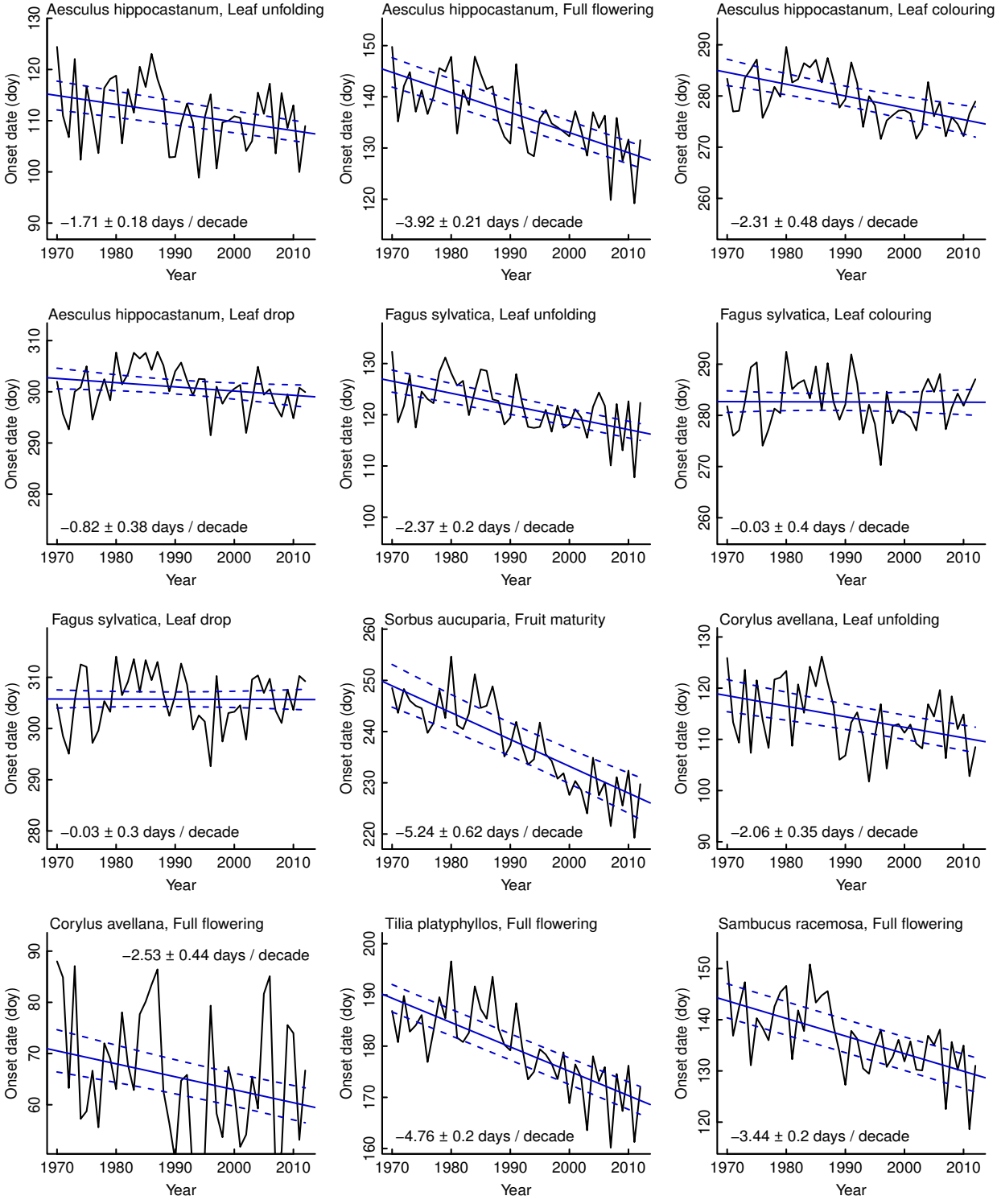
Appendix 4 (cont.): Regression of mean onset dates (1996–2012) against altitude.



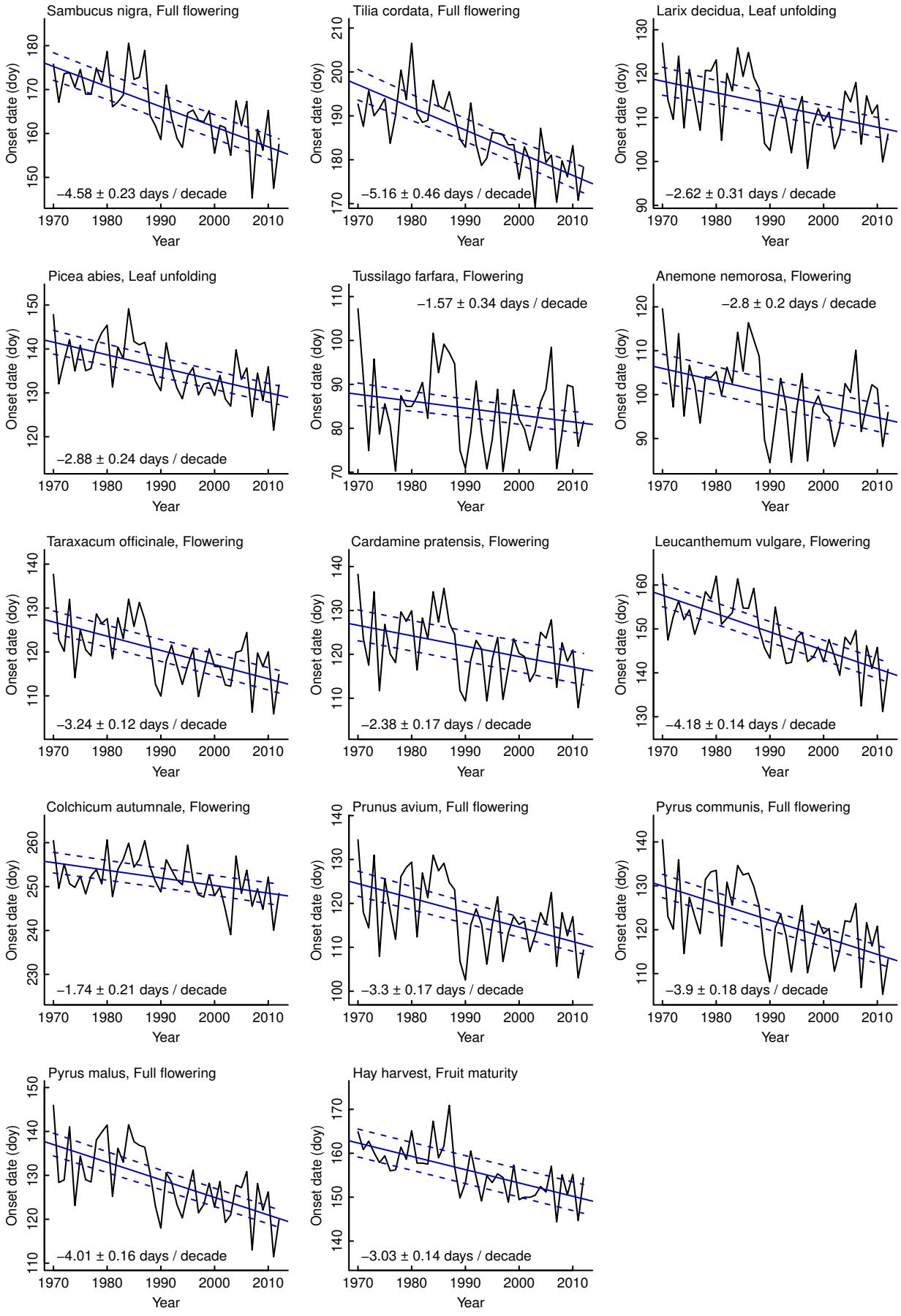
Appendix 5: Distribution of deviations from the overall altitudinal trend among phenophases for each station, based on mean onset dates for 1996–2012.



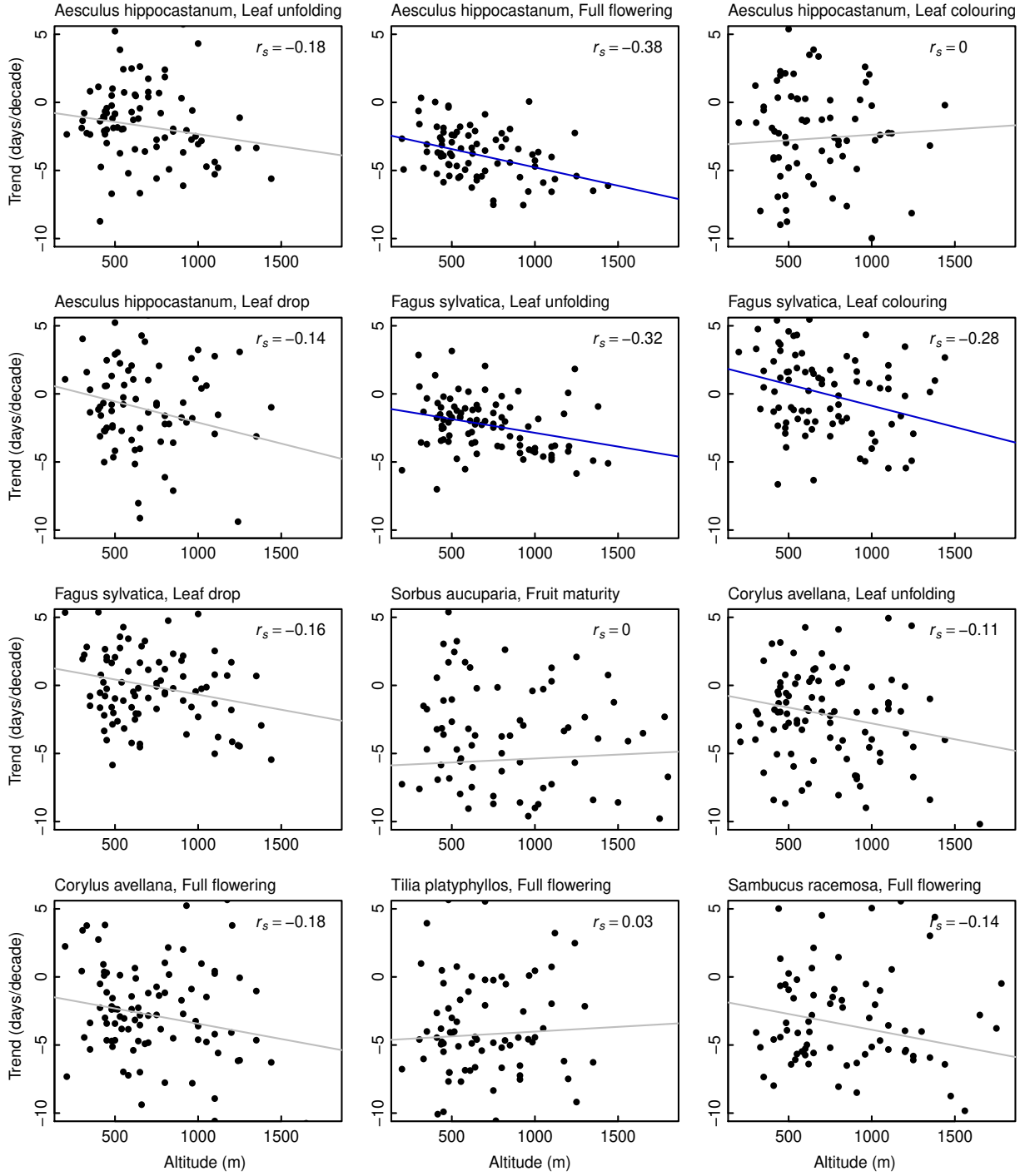
Appendix 6: Long-term year-to-year changes in phenology and linear trend (1970–2012) for each phenophase at all stations. Regression lines (with 95% confidence intervals) are derived from mixed models including random effects of stations. Regression slopes ± se are given. Negative slopes indicate a trend towards earlier onset dates. A negative trend can be regarded as statistically significant ($p < 0.05$) if mean+1.96·se < 0.



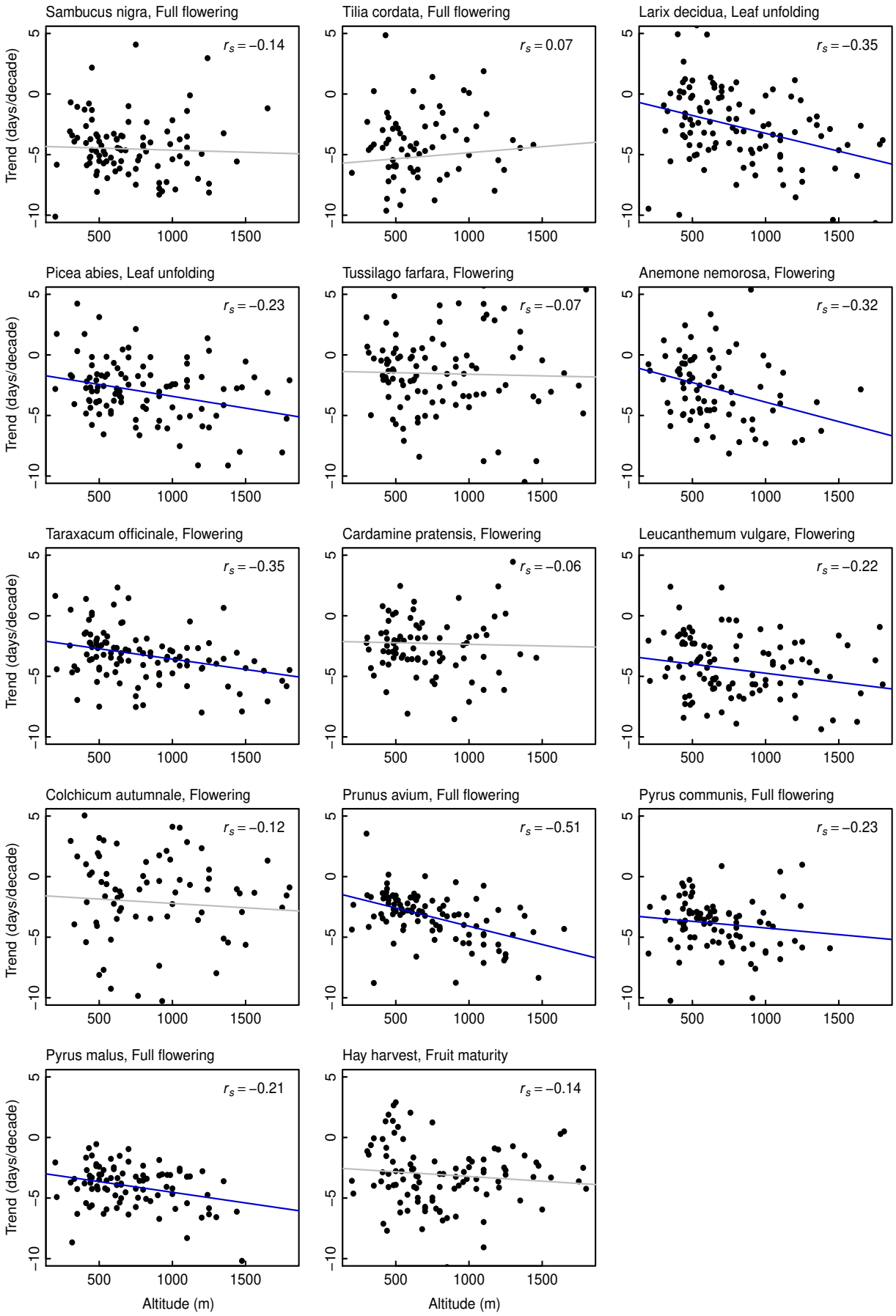
Appendix 6 (cont.): Long-term year-to-year changes in phenology and linear trend (1970-2012).



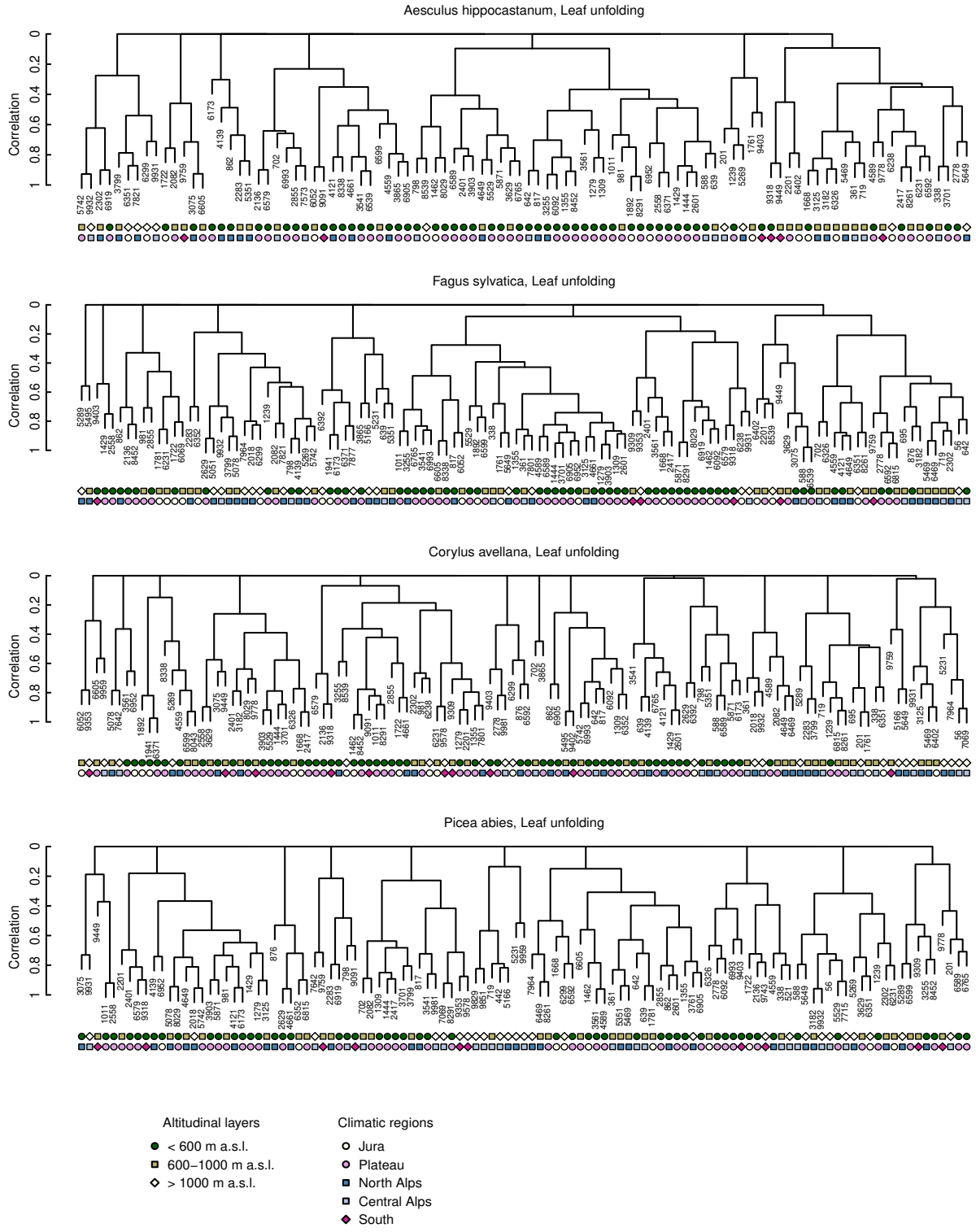
Appendix 7: Relationship between long-term trends (1970-2012) and altitude for all phenophases. Each dot represents one station. Regression lines are blue if the relationship is statistically significant ($p < 0.05$) and grey otherwise.



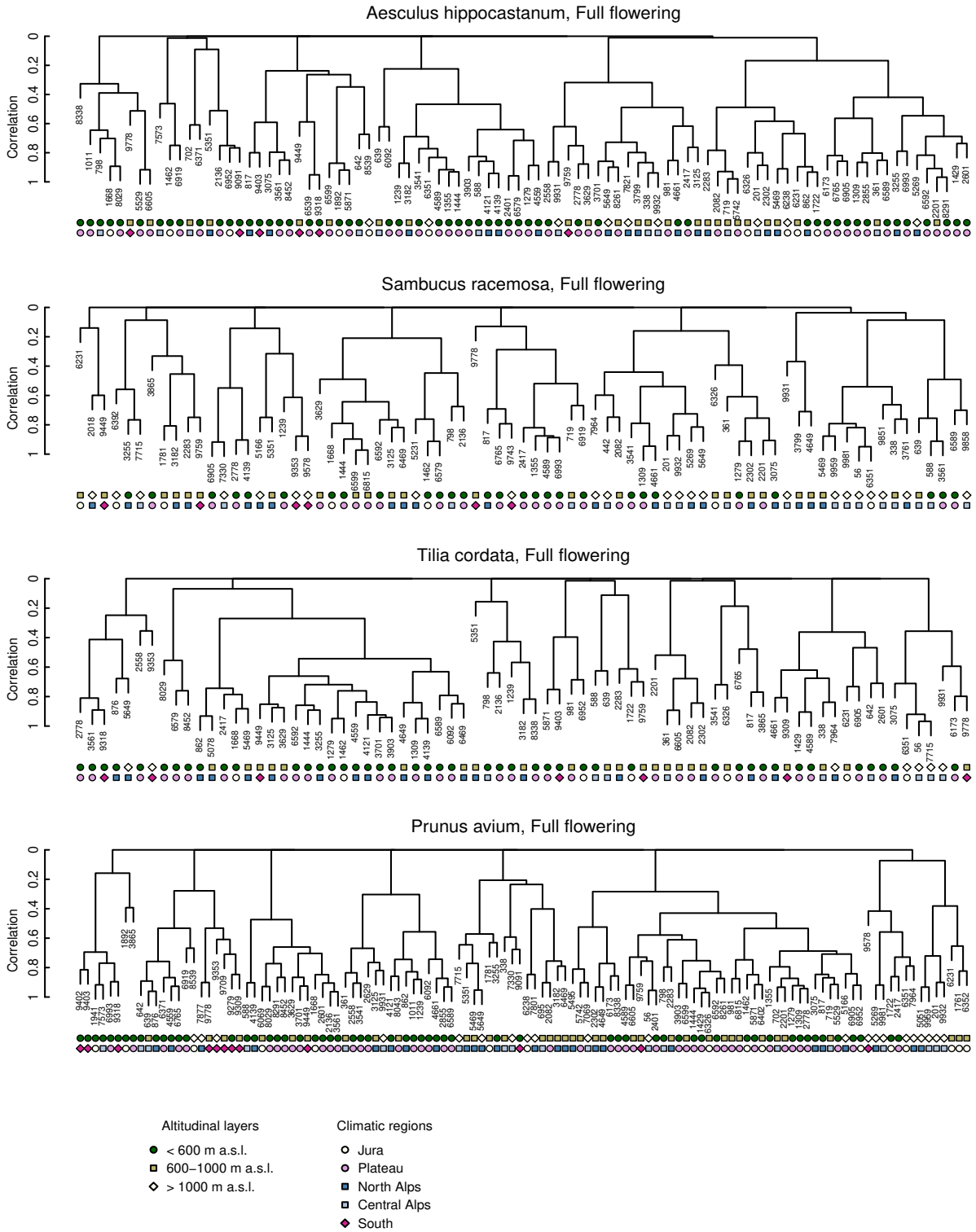
Appendix 7 (cont.): Relationship between long-term trends (1970-2012) and altitude.



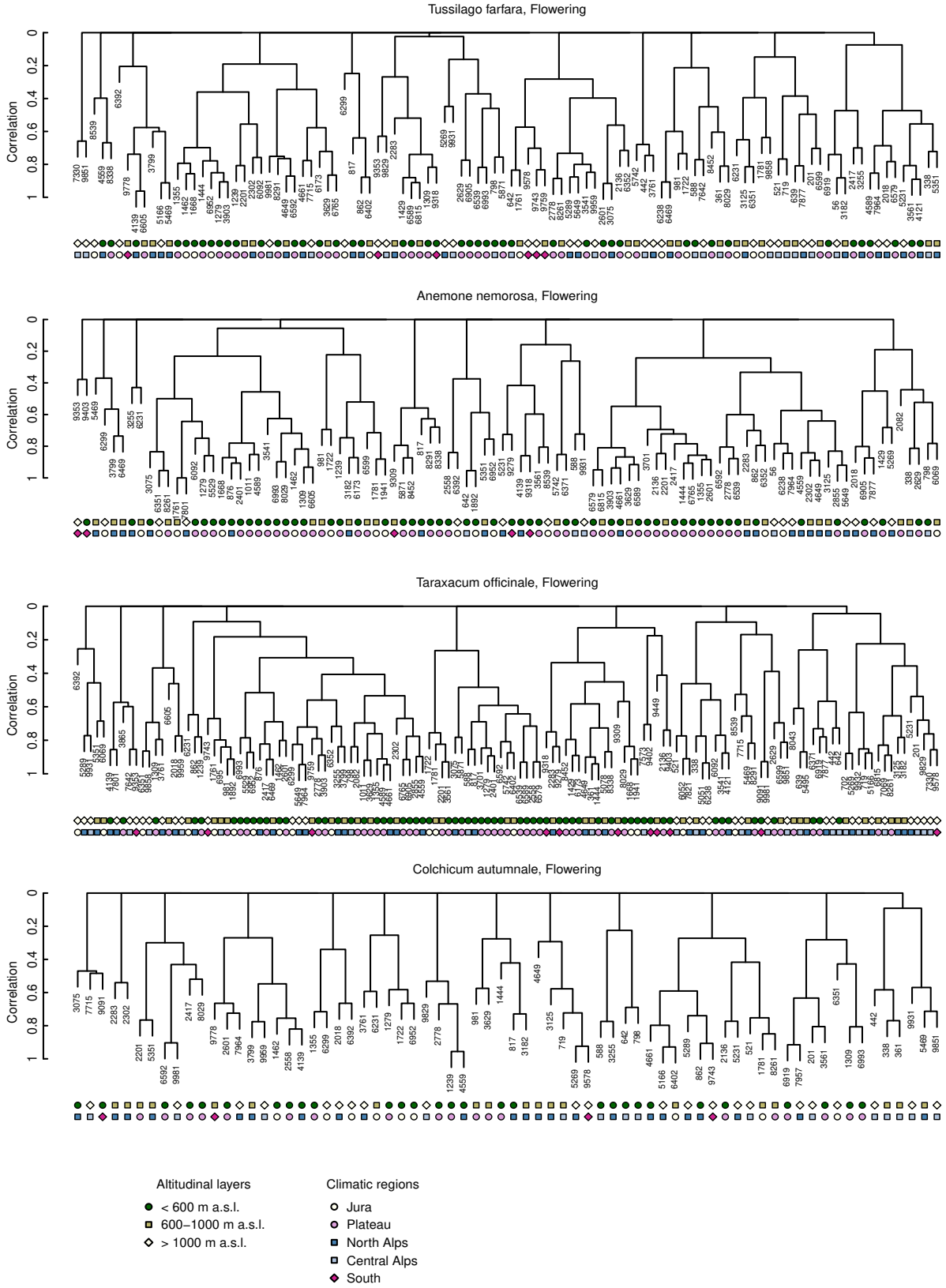
Appendix 8: Clustering of stations: dendrograms resulting from hierarchical clustering with complete linkage based on correlations between the time series of stations (1996–2012) for four selected phenophases per phenological stage. Symbols below dendrograms indicate the altitude and region of stations.



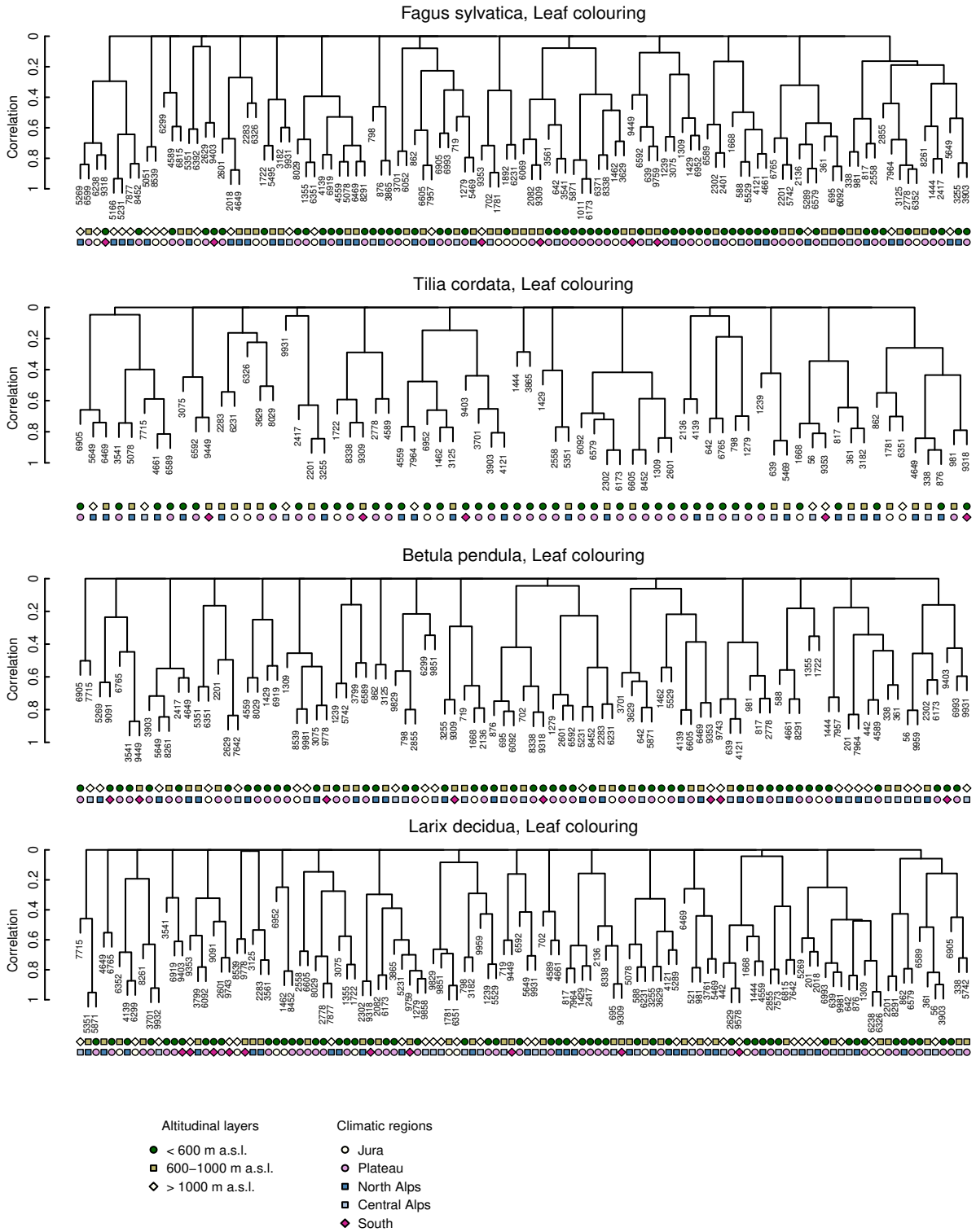
Appendix 8 (cont.): Clustering of stations based on correlations between time series (1996–2012).



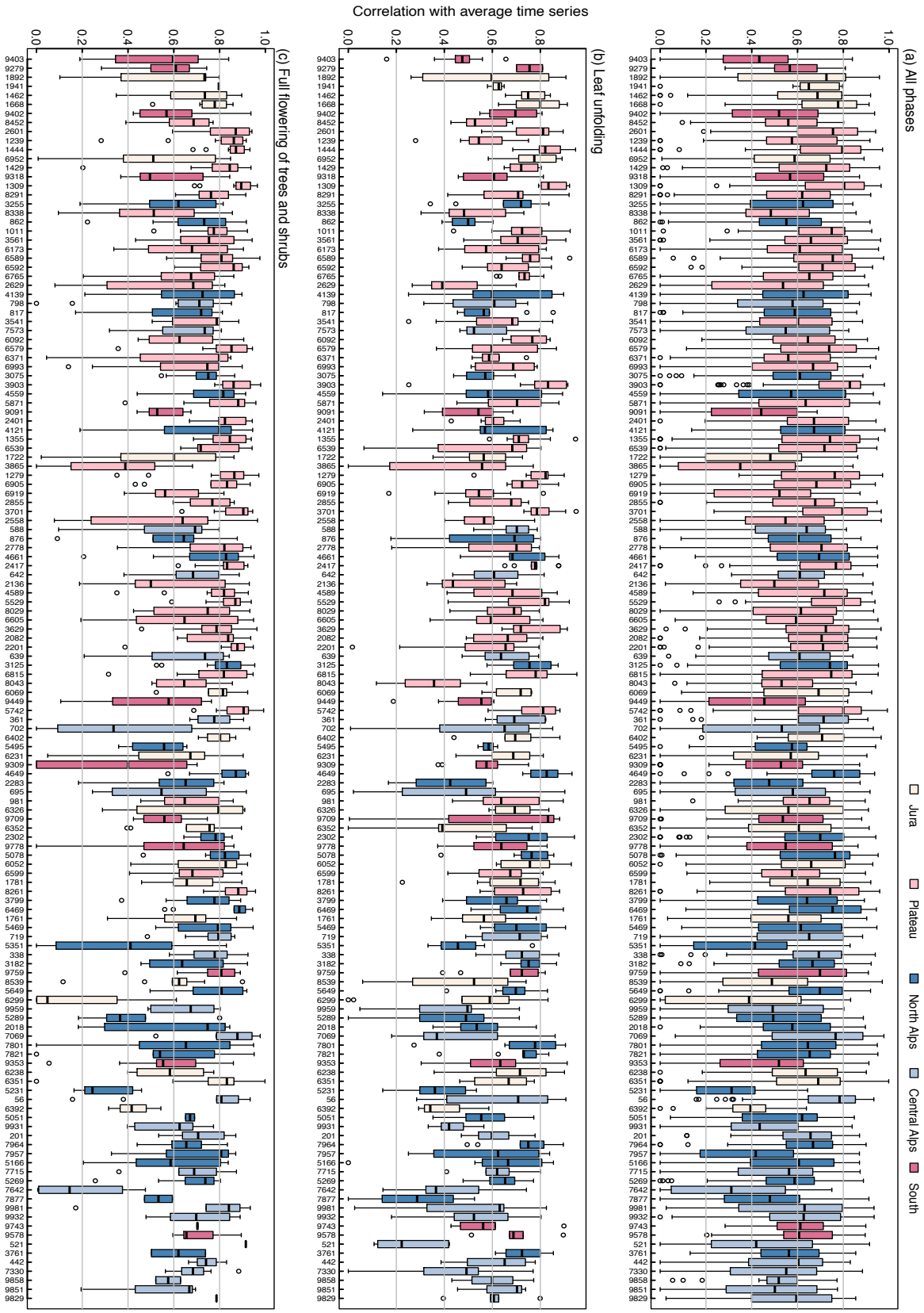
Appendix 8 (cont.): Clustering of stations based on correlations between time series (1996–2012).



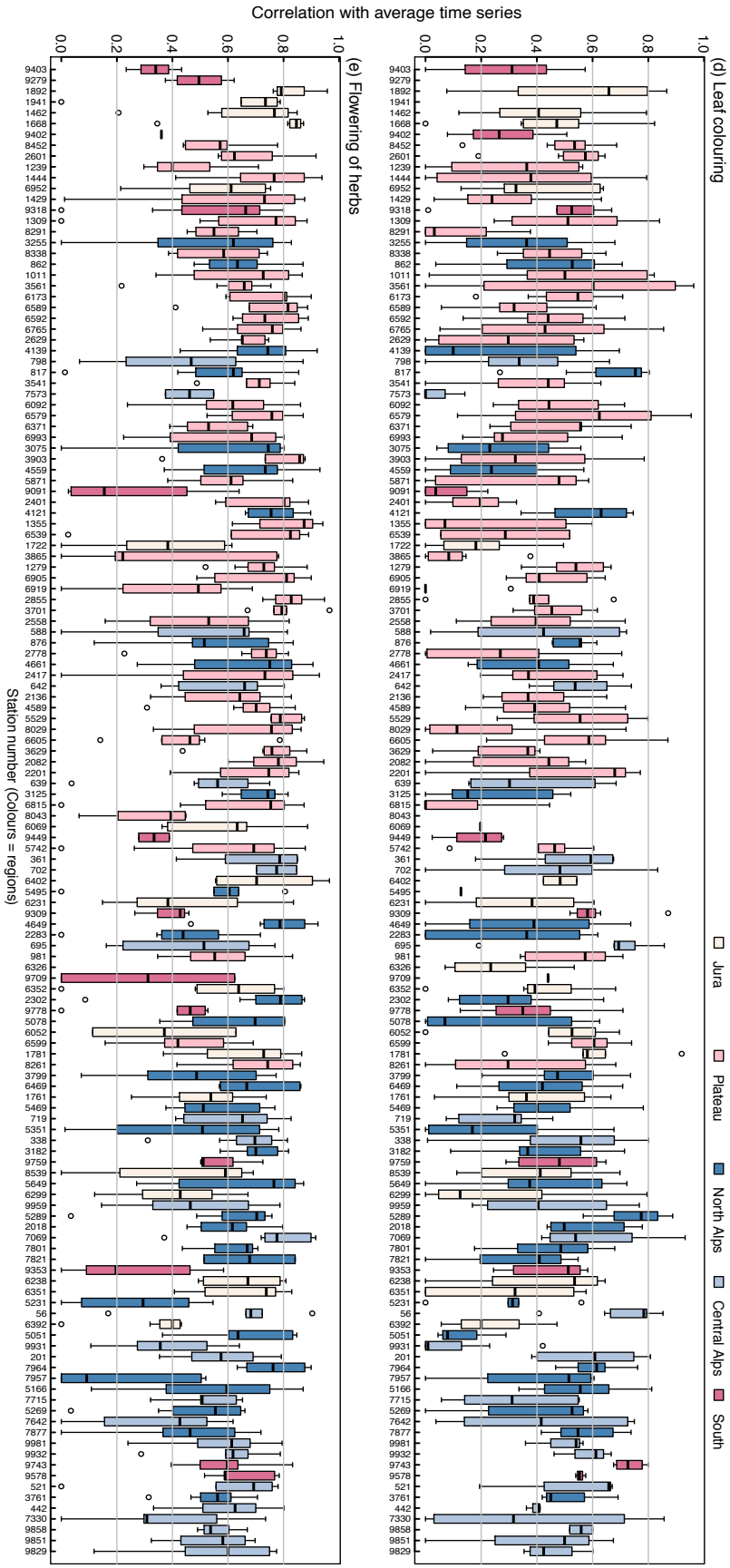
Appendix 8 (cont.): Clustering of stations based on correlations between time series (1996–2012).



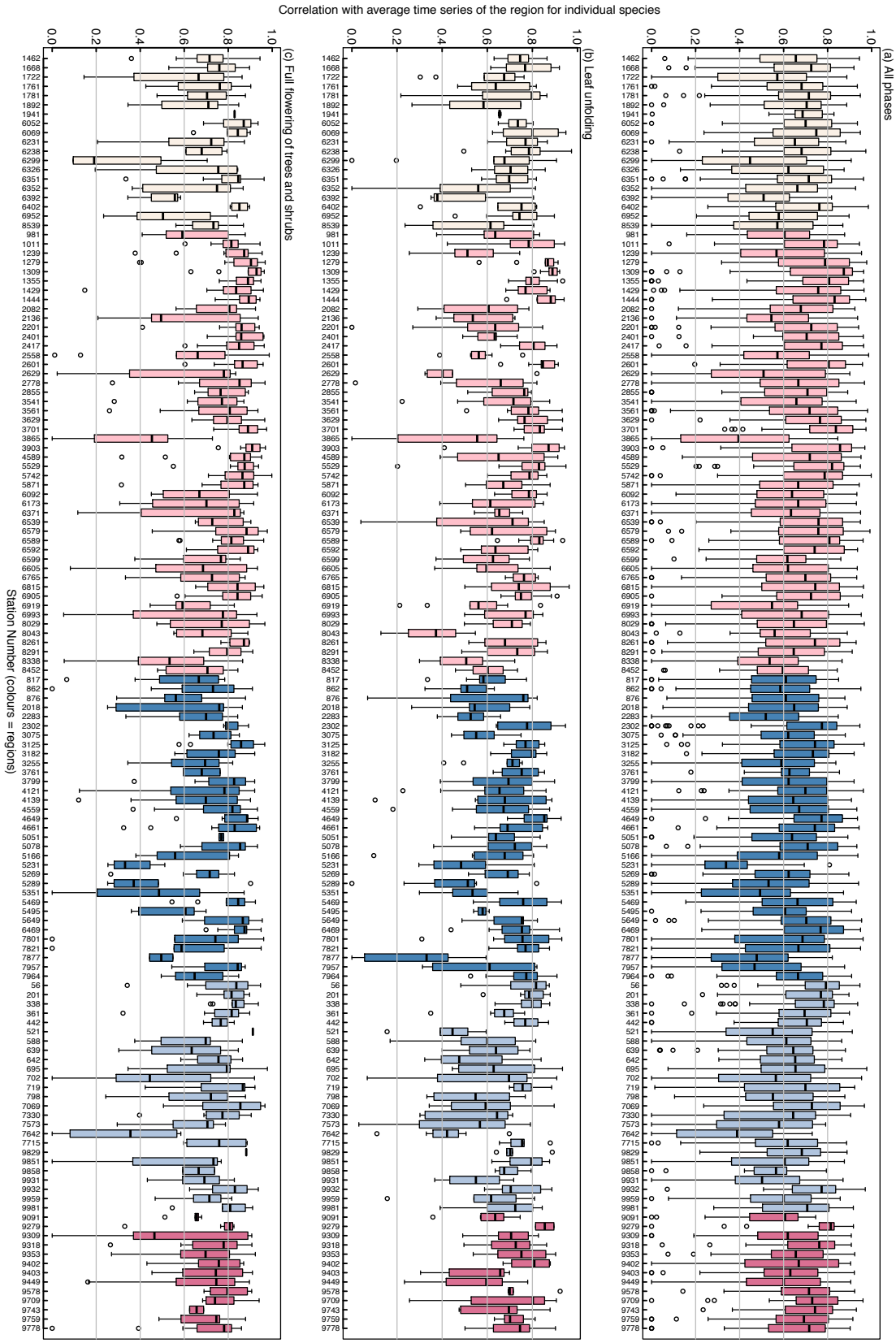
Appendix 9A: Distribution of correlations between the time series (1996–2012) of individual stations and the mean time series of all stations as a measure of each station's representativeness. Correlations were calculated separately for each phenophase; distributions of these correlations (a) over all phenophases and (b–e) over 8–11 species per phenological stage are represented by boxplots. Stations are ordered by increasing altitude from left to right, and colours indicate climatic regions.



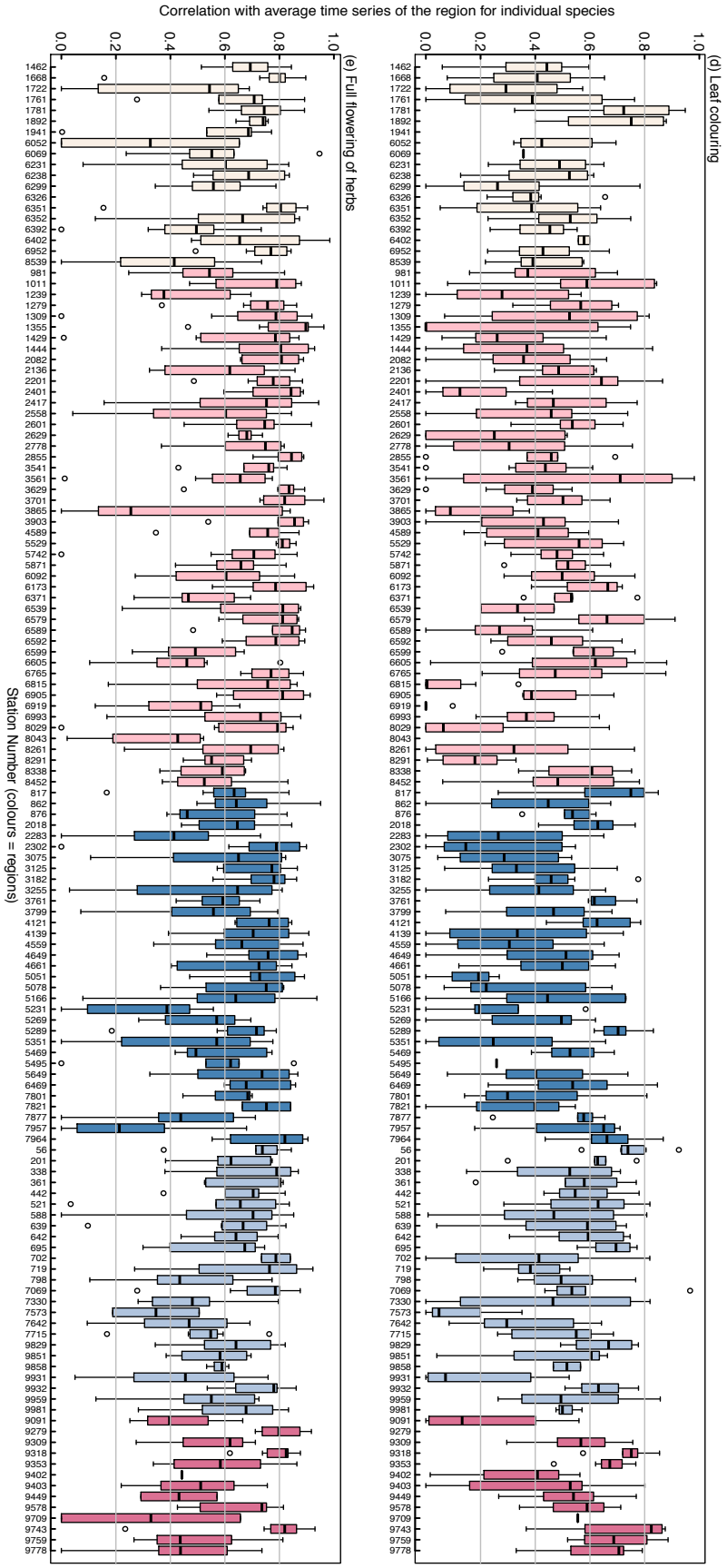
Appendix 9A (cont.): Distribution of correlations between the time series (1996–2012) of individual stations and the mean time series of all stations.



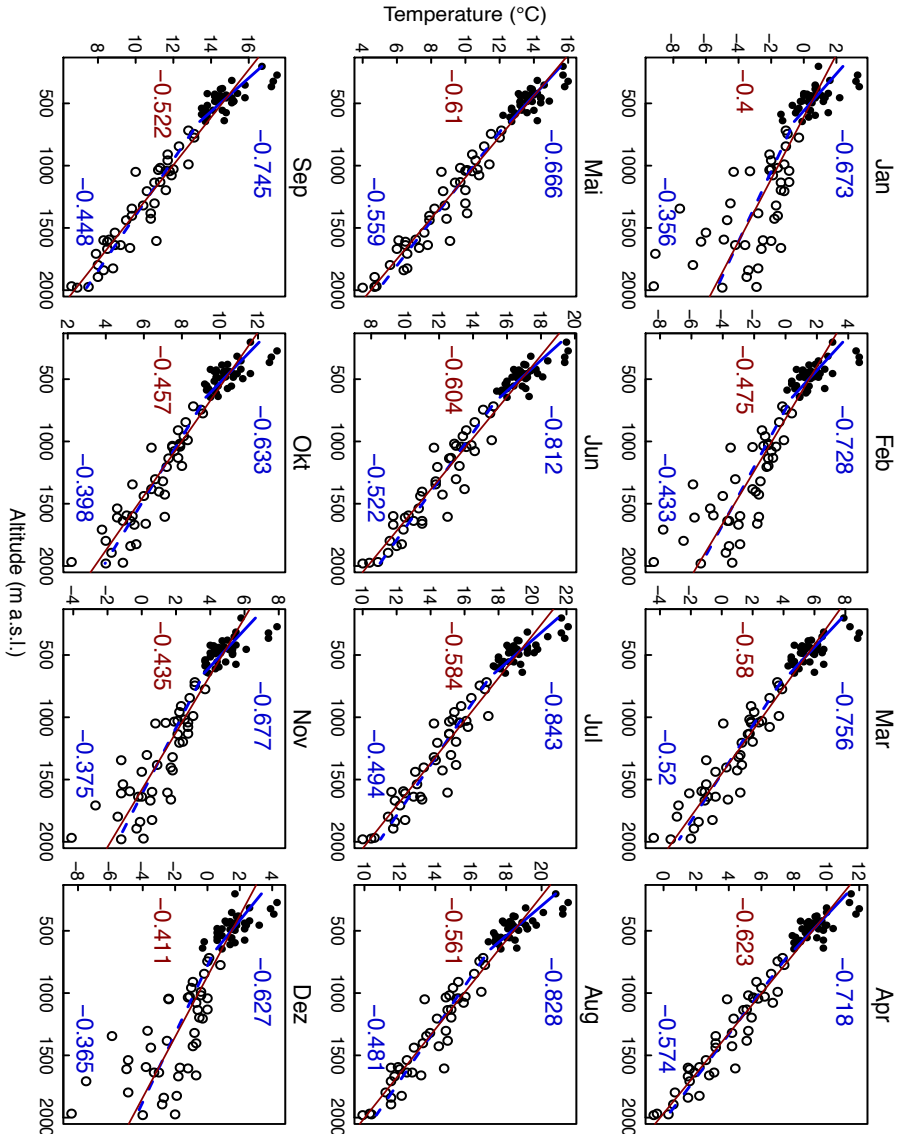
Appendix 9B: Distribution of correlations between the time series (1996–2012) of individual stations and the mean time series of each climatic region. Correlations were calculated separately for each phenophase; distributions of these correlations (a) over all phenophases and (b–e) over 8–11 species per phenological stage are represented by boxplots. Stations are ordered by climatic regions.



Appendix 9B (cont.): Distribution of correlations between the time series (1996–2012) of individual stations and the mean time series of each climatic region.



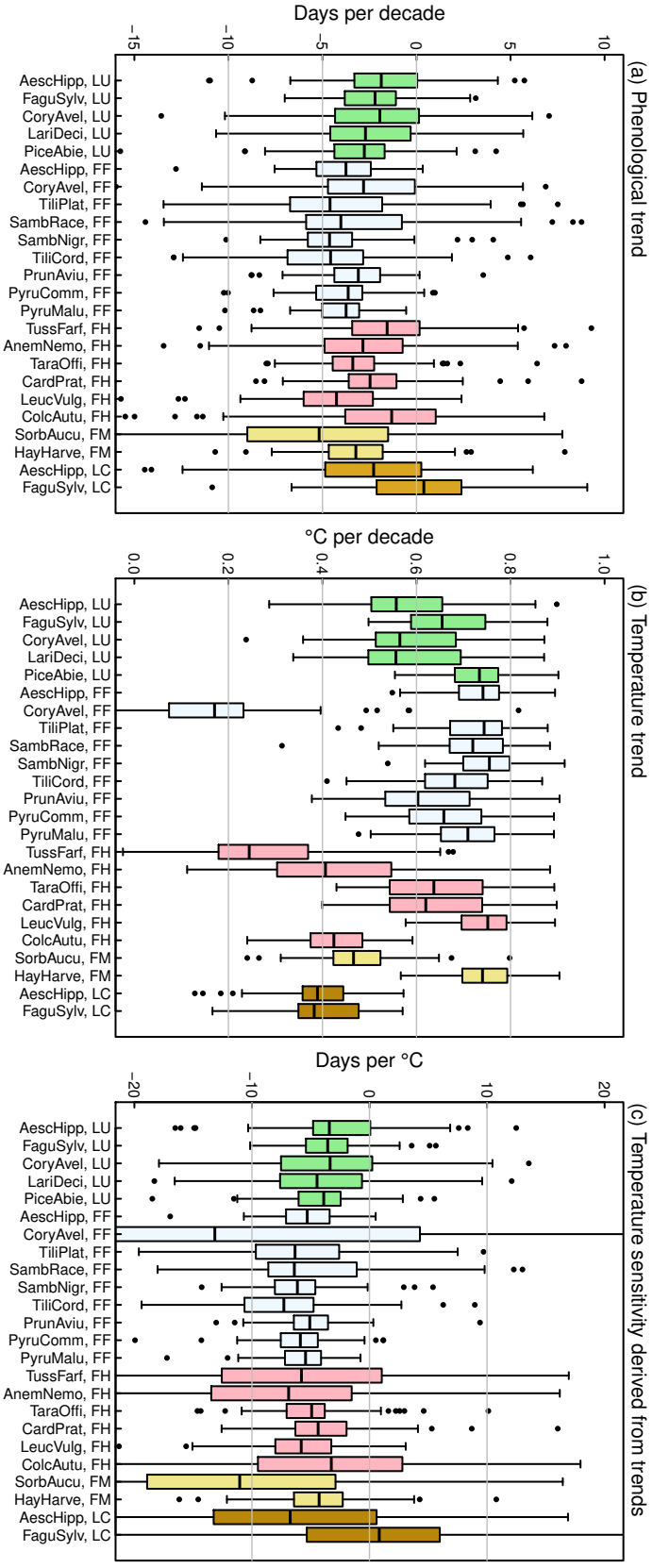
Appendix 10: Altitudinal gradients in temperature, based on norm values 1981–2010 for monthly mean temperatures at 89 climatic stations. Regression lines and slope coefficients are given in red for all stations, and in blue separately for stations above and below 700 m a.s.l.



Appendix 11: Temperature sensitivity (days/°C) of each phenophase, derived from four types of spatial or temporal variation in phenology and temperature.

Phenophase	altitudinal gradients	fluctuations 1996–2012	fluctuations 1970–2012	linear trend 1970–2012
Aesculus hippocastanum, Leaf unfolding	-5.08 (0.43)	-3.61 (0.14)	-4.05 (-4.05)	-2.83 (0.57)
Aesculus hippocastanum, Flowering start	-5.04 (0.30)	-4.60 (0.16)		
Aesculus hippocastanum, Full flowering	-5.01 (0.27)	-5.02 (0.16)	-4.94 (-4.94)	-5.33 (0.32)
Aesculus hippocastanum, Leaf colouring	-0.62 (0.73)	0.88 (0.38)	1.45 (1.45)	-7.79 (1.74)
Aesculus hippocastanum, Leaf drop	-0.22 (0.57)			
Fagus sylvatica, Leaf unfolding	-3.21 (0.24)	-3.39 (0.11)	-2.73 (-2.73)	-3.34 (0.30)
Fagus sylvatica, Leaf colouring	1.84 (0.54)	3.17 (0.26)	2.95 (2.95)	-0.31 (1.16)
Fagus sylvatica, Leaf drop	1.89 (0.41)			
Acer pseudoplatanus, Leaf unfolding	-4.04 (0.24)	-3.81 (0.14)		
Acer pseudoplatanus, Leaf colouring	1.70 (0.39)	2.81 (0.25)		
Sorbus aucuparia, Leaf unfolding	-4.37 (0.30)	-3.81 (0.15)		
Sorbus aucuparia, Flowering start	-4.48 (0.27)	-4.16 (0.16)		
Sorbus aucuparia, Full flowering	-4.97 (0.29)	-3.95 (0.38)		
Sorbus aucuparia, Fruit maturity	-2.87 (0.66)	-0.68 (0.56)	-0.99 (-0.99)	-11.47 (1.42)
Sorbus aucuparia, Leaf colouring	0.13 (0.43)	1.82 (0.25)		
Sorbus aucuparia, Leaf drop	0.32 (0.39)			
Corylus avellana, Leaf unfolding	-4.68 (0.35)	-3.54 (0.17)	-4.47 (-4.47)	-3.45 (0.59)
Corylus avellana, Flowering start	-6.20 (0.49)	-8.04 (0.34)		
Corylus avellana, Full flowering	-5.87 (0.52)	-6.83 (0.28)	-6.54 (-6.54)	-6.97 (7.53)
Tilia platyphyllos, Leaf unfolding	-4.14 (0.33)	-4.22 (0.16)		
Tilia platyphyllos, Flowering start	-4.41 (0.66)	-5.01 (0.35)		
Tilia platyphyllos, Full flowering	-5.14 (0.54)	-4.70 (0.35)	-4.76 (-4.76)	-6.01 (0.69)
Tilia platyphyllos, Leaf colouring	1.38 (0.64)	2.60 (0.33)		
Sambucus racemosa, Flowering start	-4.81 (0.60)	-4.94 (0.33)		
Sambucus racemosa, Full flowering	-4.93 (0.57)	-4.97 (0.33)	-4.67 (-4.67)	-4.88 (0.77)
Sambucus racemosa, Fruit maturity	-5.14 (0.92)	-3.34 (0.66)		
Sambucus nigra, Flowering start	-6.00 (0.41)	-5.62 (0.24)		
Sambucus nigra, Full flowering	-6.11 (0.38)	-5.44 (0.29)	-4.12 (-4.12)	-6.09 (0.39)
Sambucus nigra, Fruit maturity	-4.92 (0.61)	-1.10 (0.42)		
Tilia cordata, Leaf unfolding	-3.59 (0.37)	-4.15 (0.19)		
Tilia cordata, Flowering start	-4.64 (0.84)	-4.89 (0.34)		
Tilia cordata, Full flowering	-4.59 (0.73)	-5.21 (0.37)	-4.67 (-4.67)	-7.65 (0.67)
Tilia cordata, Leaf colouring	0.17 (0.60)	3.06 (0.32)		
Larix decidua, Leaf unfolding	-4.23 (0.26)	-4.08 (0.15)	-5.07 (-5.07)	-4.34 (0.50)
Larix decidua, Leaf colouring	1.73 (0.33)	2.49 (0.23)		
Larix decidua, Leaf drop	1.58 (0.37)			
Picea abies, Leaf unfolding	-4.22 (0.28)	-3.90 (0.18)	-4.50 (-4.50)	-4.11 (0.35)
Betula pendula, Leaf unfolding	-4.33 (0.28)	-3.87 (0.14)		
Betula pendula, Flowering start	-4.99 (0.63)	-2.91 (0.28)		
Betula pendula, Full flowering	-5.00 (0.68)	-3.18 (0.32)		
Betula pendula, Leaf colouring	1.54 (0.40)	2.14 (0.31)		
Betula pendula, Leaf drop	1.56 (0.39)			
Tussilago farfara, Flowering herbs	-3.62 (0.34)	-3.62 (0.16)	-4.86 (-4.86)	-5.67 (2.40)
Anemone nemorosa, Flowering herbs	-5.61 (0.54)	-3.54 (0.19)	-4.40 (-4.40)	-7.40 (1.47)
Dactylis glomerata, Flowering herbs	-5.41 (0.54)	-4.39 (0.30)		
Taraxacum officinale, Flowering herbs	-4.98 (0.28)	-4.18 (0.16)	-4.66 (-4.66)	-5.03 (0.35)
Epilobium angustifolium, Flowering herbs	-3.41 (1.18)	-2.86 (0.62)		
Cardamine pratensis, Flowering herbs	-6.64 (0.46)	-4.61 (0.25)	-4.20 (-4.20)	-3.72 (0.49)
Leucanthemum vulgare, Flowering herbs	-4.81 (0.28)	-3.91 (0.23)	-4.60 (-4.60)	-5.97 (0.38)
Colchicum autumnale, Flowering herbs	-0.65 (0.53)	-1.39 (0.36)	-0.17 (-0.17)	-5.12 (1.29)
Prunus avium, Flowering start	-4.90 (0.29)	-4.27 (0.12)		
Prunus avium, Full flowering	-5.17 (0.27)	-4.52 (0.12)	-5.64 (-5.64)	-5.12 (0.29)
Pyrus communis, Flowering start	-4.53 (0.39)	-5.10 (0.14)		
Pyrus communis, Full flowering	-4.68 (0.33)	-5.27 (0.13)	-5.94 (-5.94)	-5.94 (0.34)
Pyrus malus, Flowering start	-4.29 (0.31)	-4.83 (0.14)		
Pyrus malus, Full flowering	-4.54 (0.27)	-5.02 (0.13)	-5.64 (-5.64)	-5.73 (0.25)
Hay harvest, Fruit maturity	-6.70 (0.41)	-2.64 (0.30)	-3.15 (-3.15)	-4.15 (0.38)

Appendix 12: Distribution of temperature sensitivity coefficients derived from linear trends between 1970 and 2012. (a) Distribution of phenological trends (slopes of linear regression against years), (b) distribution of the associated temperature trends, and (c) distribution of sensitivity coefficients calculated as the ratio between phenological trends and temperature trends. All coefficients were determined for each phenophase and station; boxes represents distributions over stations for each phenophase.



MeteoSchweiz
Operation Center 1
CH-8044 Zürich-Flughafen
T +41 58 460 99 99
www.meteoschweiz.ch

MeteoSvizzera
Via ai Monti 146
CH-6605 Locarno Monti
T +41 58 460 97 77
www.meteosvizzera.ch

MétéoSuisse
7bis, av. de la Paix
CH-1211 Genève 2
T +41 58 460 98 88
www.meteosuisse.ch

MétéoSuisse
Chemin de l'Aérologie
CH-1530 Payerne
T +41 58 460 94 44
www.meteosuisse.ch